Error Estimates of Best R-Approximation in TT-Tensor Format

Fabian Maximilian Faulstich
Technische UniversitÃt Berlin

Supervisor: Prof. Dr. Reinhold Schneider

A thesis submitted for the degree of

Bachelor of Science

in Mathematics

Statement of Authorship

This thesis has been submitted for the degree of Bachelor of Science in Mathematics. I, the undersigned, hereby declare that:

- I am the sole author of this thesis.
- I have fully acknowledged and referenced the ideas and work of others, whether published or unpublished, in my thesis.
- I have prepared my thesis specifically for the degree of Bachelor of Science while under supervision at the Technical University of Berlin.
- My thesis does not contain work extracted from a thesis, dissertation or research paper previously presented for another degree or diploma at this or any other university.

Berlin, the 13th of August 2015: _____

Fabian Faulstich

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfmittel angefertigt habe.

Berlin, den 13.08.2015: _____

Fabian Faulstich

Abstract

In this thesis an error estimate for the residual of the Krylov subspace methods for r-rank tensor train approximations of physical systems using the nearest neighbour interaction approximation is presented. We start by giving a summary of tensors, tensor networks, the tensor train format and the ALS algorithm. Furthermore we introduce the first and second quantisation, spin chains, solid state physics and the Hubbard model. These mathematical and physical structures will be combined to simulate a physical system. This motivates the analytical part of the thesis. Our approach to the final error estimate uses the Krylov subspace methods where for each iteration step a rank estimate for the residual is given. Furthermore Chebyshev polynomials will be used to derive an optimal bound for Krylov subspace methods. To further improve this bound for the special case of a physical system a Hamiltonian of a one-dimensional spin chain constrained by the nearest neighbour interaction will be considered. These conditions ameliorate the properties of the operator, which will improve the bound. Additionally we prove a statement about the approximability of a physical system in dependence of the Renyi entropy. The approximability of a physical system by the tensor train format is a necessary condition for the validity of the bound.

Zusammenfassung

Im Folgenden erarbeiten wir eine Fehlerabschätzung des Residuums von Krylov-Unterraum-Verfahren, welche eine Rang-r-Tensor-Train-Approximation eines physikalischen Systems berechnen, wobei die Nächste-Nachbar-Wechselwirkung zur Approximation verwendet wird. Wir beginnen mit einer Zusammenfassung über Tensoren, Tensornetzwerke, das Tensor-Train-Format und den ALS-Algorithmus. Weiter geben wir eine Einführung in die erste und zweite Quantisierung, Spin-Ketten, Festkörperphysik und das Hubbard-Modell. Diese mathematischen und physikalischen Strukturen werden kombiniert, um ein physikalisches System zu simulieren. Dies motiviert den analytischen Teil der Arbeit. Die abschließende Fehlerabschätzung nutzt die Krylov-Unterraum-Verfahren, wobei in jeder Iteration eine Rangabschätzung des Residuums vorgenommen wird. Weiter werden Tschebyscheff-Polynome verwendet, um die Optimalität der Schranke für Krylov-Unterraum-Verfahren sicherzustellen. Um diese Schranke für den Spezialfall eines physikalischen Systems zu verbessern, betrachten wir einen Hamiltonoperator einer eindimensionalen Spin-Kette, welche der Nächste-Nachbar-Wechselwirkung unterliegt. Diese Einschränkungen verleihen dem Operator zusätzliche Eigenschaften, welche zu einer Verbesserung der Schranke führen. Des Weiteren treffen wir eine Aussage über die Approximierbarkeit eines physikalischen Systems in Abhängigkeit von der Renyi-Entropie. Die Approximierbarkeit eines physikalischen Systems durch das Tensor-Train-Format ist eine notwendige Voraussetzung für die Gültigkeit der angegebenen Schranke.

Acknowledgements

I would like to thank professor Reinhold Schneider and his study group for welcoming me as student and making this thesis possible. In particular I would like to thank Benjamin Huber for his guidance, time and patience during our weekly meetings, even in the hot summertime. Thomas Jankuhn and Sebastian Zachrau for fruitful discussions and objective questioning the given statements. I owe particular thanks to Mathieu Rosière for rereading this thesis, finding mathematical weak points and the effort he made to familiarise himself with this subject.

Contents

1 Introduction

In 1927 a collection of information was published by Niels Bohr and Werner Heisenberg that postulated how the mathematical formalism of quantum mechanics is supposed to be understood in terms of everyday language [1]. This publication gave a unique interpretation of the analysed quantities. This so-called *Copenhagen interpretation* was the start of a period full of new theories and models which aimed to describe physical systems. Since then there has been huge progress in the field of theoretical physics especially in quantum theory, which made it possible to quantise physical systems which became more and more complex.
As nice as these formulations are the numerical simulation of such systems becomes increasingly difficult as the storage scales exponentially in the dimension. This is where new numerical theories become of interest to predict the behaviour of such systems.
One big step in simulating large systems was the discovery of special tensor formats like the *tensor train format*. Tensors appear naturally in *many body quantum systems* as objects that describe the state of a system. The scaling behaviour of a tensor suffers from the curse of dimensionality, which means that it grows exponentially with the dimension. This curse was broken with the discovery of the *tensor train format*. This approximation format and the optimisation algorithms therewith connected like the *ALS* have been used for quite a long time in *quantum many body physics*. It led to a basic understanding of *quantum many body systems* and showed that predictions made by physical models were accurate.

The aim of this thesis is to calculate an error estimate for the *tensor equation* $Ax = b$ where A is a *tensor operator* and $x,\ b$ are *tensors*. These objects will be considered in the *tensor train format*. To fulfil this aim we start by summarising the basic ideas of *quantum many body systems* and the *tensors train format*. Furthermore they will be combined to calculate a bound for the error made by *Krylov subspace methods*.
For numerical simulations the *Hubbard model* and the *tensor train format* are used to get experimental results about the behaviour of the residual. These results shall be compared to the analytical error bound.
We start this thesis with a chapter introducing the *tensor-algebra*, the *tensor train format* and the *alternating least squares (ALS) algorithm*. The second chapter gives an introduction to the physical background used in this thesis such as *first* and *second quantisation*, *the Hubbard model* and *spin chains*. This is followed by a chapter that presents numerical results using the *tensor train format* and *the Hubbard model* to

approximate the eigenvectors of the Hamiltonian. After these experimental results a theoretical discussion of the residual in dependence of the *tensor train rank* is given. The last chapter will focus on the question of whether a physical system can be approximated in the *tensor train format*.

2 Tensor Structures

Just as an $n \times m$-matrix can be described as a system of numbers $(a_{i,j})$ with the indices $i \in \{1, .., n\}$ and $j \in \{1, ..., m\}$ higher dimensional objects can be defined. These objects are a natural generalisation of a matrix characterised by an arbitrary but finite number of numbers $a_{i_1,...,i_d}$ where $i_j \in \{1, ..., n_j\}$ holds for each index. Such objects arise naturally when higher dimensional systems are considered. One example is *quantum many body physics* where the quantum mechanical states are described by such object.
This chapter gives a brief introduction to these so-called *tensors*. It is important to notice that only the prerequisites necessary for this thesis are mentioned. For more detailed information see [2], [3].
This chapter starts with the fundamental structures considered in this thesis the *tensors* and *tensor networks*. Further the *diagrammatic notation* is introduced. If high dimensional objects like tensors are used for numerical purposes, they demand a lot of storage. This behaviour can be described by the curse on dimensionality which states that the storage of a tensor scales exponentially in the dimension. Therefore the *tensor train format*, which approximates *tensors* and breaks this curse of dimensionality, is introduced. As *tensors* appear in optimisation problems, a treatment in the *tensor train format* is of need. The *ALS* algorithm is a way of doing this and is introduced at the end of this chapter.

2.1 Tensors and Tensor Networks

We start this section with the definition of a *tensor*, which will be the fundamental object in this thesis.

Definition: A map $T : \Omega_1 \times \Omega_2 \times ...\tilde{A}^{-}\Omega_d \to \mathbb{R}$ with $\Omega_j = \{1, 2, ..., n_j\}$ is called a *tensor of order (or degree) d* and *dimensions* $n_j = |\Omega_j|$. The elements $i \in \Omega_j$ are called *indices* and a tuple thereof $\mathbf{i} = (i_1,\ i_2, ...,\ i_d) \in \Omega_1 \times ... \times \Omega_d$ a *d-dimensional multi-index*. The set of all tensors with the canonical addition form a vector space isomorphic to (and often identified with) $\mathbb{R}^{\mathbf{n}} = \mathbb{R}^{n_1 \cdot ... \cdot n_d} = \mathbb{R}^{n_1} \times ... \times \mathbb{R}^{n_d}$. The degenerate case of an order-0 tensor is called a scalar and can canonically be identified with an element of $\mathbb{R}$. In the following $\mathbb{T}(d, \mathbf{n})$ describes the space of all order d tensors where $\mathbf{n} = (n_1, ..., n_d)$. Furthermore

$$\mathbb{T}(d) := \bigcup_{\mathbf{i} \in \mathbb{N}^d} \mathbb{T}(d, \mathbf{i}) \ .$$

It is important to notice that unlike $\mathbb{T}(d, \mathbf{n})$ the set $\mathbb{T}(d)$ is not a linear space. As *tensor networks*, *tensor subnetworks* or *TT tensors* are introduced later the term *full tensor* is used to distinguish those from a tensor as defined above.
Tensors are a generalisation of matrices but it is unclear how to understand the inversion or the SVD of these objects. As the inversion or the SVD is well understood for matrices, a bijection between tensors and matrices has to be found. Such a bijection is called *matricisation*. A *matricisation* uses so-called *flattenings* which are bijections between tensor spaces of different order tensors.

Definition: Let $\Omega_1, ..., \Omega_n$ be sets of indices. A bijection

$$\Lambda : \Omega_1 \times ... \times \Omega_k \to \Omega_\Lambda = \{1, ..., n_\Lambda\} \qquad \text{, where } n_\Lambda = \prod_{j=1}^{k} |\Omega_j|$$

is called a *flattening*. The inverse of this function is called an *expansion*.

Having defined the flattening, the *matricisation* is a natural consequence. It is possible to define two flattening acting on disjoint sets of indices which together form the entire set of indices. As the flattenings are bijections this combination of bijections is again a bijection. This is the idea of the *matricisation*.

Definition: Let T be a tensor of order d. Define two flattenings

$$\Lambda_1 : \Omega_1 \times ... \times \Omega_k \to \Omega_{\Lambda_1}$$
$$\Lambda_2 : \Omega_{k+1} \times ... \times \Omega_d \to \Omega_{\Lambda_2} \ .$$

The *matricisation* of the tensor T with respect to the index $k \leq d$ is the matrix $M_k(T)$ given by

$$M_k(T)_{i,j} = T_{\Lambda_1^{-1}(i), \Lambda_2^{-1}(j)}$$

for indices $(i, j) \in \Omega_{\Lambda_1} \times \Omega_{\Lambda_2}$. The inverse of such a matricisation is called *tensorisation*

The matricisation makes it possible to multiply two matricised tensors. This is a so-called *contraction* along certain indices. Defining a matricisation that uses a flattening summarising all indices but the i-th of the first tensor and second matricisation that uses a flattening summarising all indices but the j-th of the second tensor. This leads to two matrices that can be multiplied. The result is a matrix which can be expanded to a tensor which is the result of the *contraction* along the i-th index of the first tensor and the j-th index of the second tensor.

Definition: Let $\mathbf{n}$ and $\mathbf{m}$ be multi-indices where $n_i = m_j$. The map

$$C_{i,j} : \mathbb{T}(d_1, \mathbf{n}) \times \mathbb{T}(d_2, \mathbf{m}) \to \mathbb{T}(d_1 + d_2 - 2)$$

that maps two tensors to a third, obtained by summing over the i-th index of $T \in \mathbb{T}(d_1, \mathbf{n})$ and the j-th index of $U \in \mathbb{T}(d_2, \mathbf{m})$

$$\sum_k T_{\alpha_1,\dots,\alpha_{i-1},k,\alpha_{i+1},\dots,\alpha_{d_1}} \cdot U_{\beta_1,\dots,\beta_{j-1},k,\beta_{j+1},\dots,\beta_{d_2}} =: R_{\alpha_1,\dots,\alpha_{i-1},\alpha_{i+1},\dots,\alpha_{d_1},\beta_1,\dots,\beta_{j-1},\beta_{j+1},\dots,\beta_{d_2}} ,$$

is called *(i,j)-contraction* of T and U. The tensor R is the image of (T, U) under the contraction $C_{i,j}$.

If it is apparent from context which indices are contracted, this contraction will be referred to as the *contraction of the common index / indices*.

Furthermore the matricisation makes it possible to think of *tensor decompositions* as a matrix decomposition applied to every possible matricisation of a tensor i.e. *partially applied* to the tensor. One decomposition using the SVD as matrix decomposition is the so-called *tensor train format*. This will be the standard form of the tensors considered in this thesis.
Before defining the tensor decomposition we recall the definition of a matrix decomposition.

Definition: A map

$$D : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times \ell} \times \mathbb{R}^{\ell \times m},\ M \mapsto (M_1, M_2) ,$$

where $M = M_1 M_2$, is called a matrix decomposition.

Based on this definition a tensor decomposition can be defined.

Definition: Let

$$\Lambda_1 : \Omega_1 \times \dots \times \Omega_k \to \Omega_{\Lambda_1}$$
$$\Lambda_2 : \Omega_{k+1} \times \dots \times \Omega_d \to \Omega_{\Lambda_2}$$

be two flattenings, where $n_1 := |\Omega_{\Lambda_1}|$, $n_2 := |\Omega_{\Lambda_2}|$ and $n := n_1 + n_2$. These flattenings define the matricisation M_k. Given a matrix decomposition

$$D : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^{n\times \ell} \times \mathbb{R}^{\ell \times m} ,$$

we define the matricisations

$$M_k^{(1)} : \mathbb{T}(k+1, \mathbf{m}_1) \to \mathbb{R}^{n_1 \times \ell},\ T \mapsto (T_{\Lambda_1^{-1}(i),j})_{i,j}$$
$$M_k^{(2)} : \mathbb{T}((n-k)+1, \mathbf{m}_2) \to \mathbb{R}^{\ell \times n_2},\ T \mapsto (T_{i,\Lambda_1^{-1}(j)})_{i,j}\ ,$$

where $\mathbf{m}_1 = (|\Omega_1|, ..., |\Omega_k|, \ell)$ and $\mathbf{m}_2 = (\ell, |\Omega_{k+1}|..., |\Omega_d|)$. The map

$$\mathcal{D} : \mathbb{T}(d, \mathbf{m}) \to \mathbb{T}(k+1) \times \mathbb{T}((d-k)+1),\ T \mapsto ((M_k^{(1)})^{-1}(A), (M_k^{(2)})^{-1}(B))\ ,$$

where $(A, B) = D \circ M_k(T)$ is the tensor decomposition induced by D with respect to Λ_1 and Λ_2.

This definition makes it possible to decompose a tensor using a decomposition scheme for matrices. Hence it justifies the SVD used partially on the matricisations, which leads to the *tensor train format*.
Tensors do not have to appear all by themselves. Often they appear in combination with other tensors and are contracted along several indices. These complex objects are called *tensor network*.

Definition: Let $\mathfrak{T}$ be a finite set of tensors and $\mathcal{C}$ be a set of triplets (T_1, T_2, R), where $T_1,\ T_2 \in \mathfrak{T}$ and R is the result of an (i, j)-contraction of T_1 and T_2. The pair $(\mathfrak{T}, \mathcal{C})$ is called a *tensor network* if for each tensor $T_1 \in \mathfrak{T}$ there exists a tensor $T_2 \in \mathfrak{T}$ and a tensor R such that $(T_1, T_2, R) \in \mathcal{C}$. Let $\mathfrak{T}_s \subseteq \mathfrak{T}$ and $\mathcal{C}_s \subseteq \mathcal{C}$. The pair $(\mathfrak{T}_s, \mathcal{C}_s)$ is a *subnetwork* of $(\mathfrak{T}, \mathcal{C})$ if it is a tensor network itself.

Considering a large set of tensors which are going to be contracted along several different indices the notation and the sets of different indices get unmanageable. Such a system stays much clearer if it is represented by a graph. This representation is called the *diagrammatic notation*. Such a graph is noted as $G = (\mathfrak{T},\ E,\ L)$ where each tensor $T \in \mathfrak{T}$ is now interpreted as a vertex with the same degree as the respective T. There exists an edge $e \in E$ with incident vertices T_1 and T_2 if and only if there exists a triplet $(T_1, T_2, R) \in \mathcal{C}$. Multiple edges will be drawn between two vertices to emphasize that more than one index is contracted. The set of not contracted indices L is usually depicted as open-ended edges in the graph and can be interpreted as multi-indices as they depend on the vertex. The number of open-ended edges in the tensor network corresponds to the degree of the full tensor that is obtained by fully contracting the network, which means to perform all contractions.

Comparing a tensor network to a full tensor it is important to notice that not

only the dimensions of the indices are of interest but also the dimensions of the contracted edges. Therefore we will denote the indices defined before as *external indices* and the contracted indices as *internal indices*.

Definition: The tuple of dimensions of all internal indices is called *rank*.

The sequence of the tuple depends on the tensor network and has to be defined respectively. The rank of the *tensor train format* will be defined later.
During discussions on bounds of residuals it simplifies the calculation if the rank is considered homogeneously. Thus speaking of the rank of a tensor network as a number, means that the rank is considered homogeneously with the value of the maximum of the actual rank of the tensor.
The elements of the rank might be written above the edges in the graph describing the tensor network.

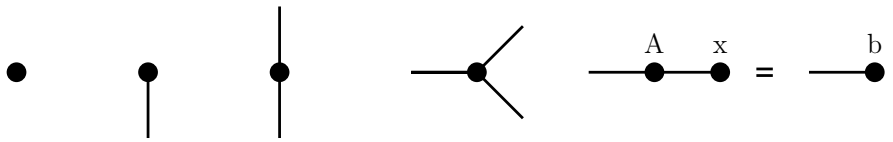
Several algorithms working with tensor networks use contractions which only act locally on a *subnetwork*. To ensure that this contraction is well-defined the following Lemma is needed.

Lemma: Any subnetwork of a tensor network is itself a tensor network and can thus be contracted to a single tensor.

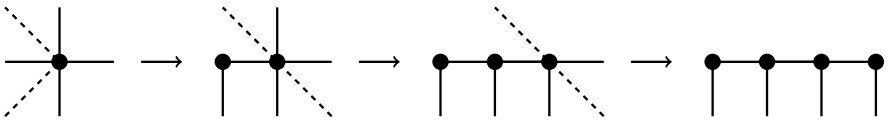
Proof: Let $(\mathfrak{T}_s, \mathcal{C}_s)$ be a subnetwork of $(\mathfrak{T}, \mathcal{C})$. We first consider those tensors of $\mathfrak{T}_s$ which are connected, i.e. the edges going from these vertices are all part of $\mathcal{C}_s$. These tensors can be contracted along the connecting edges. There might also be tensors of the subnetwork which are not connected with other tensors of the subnetwork, i.e. the edges going from this vertex are not part of the subnetwork. These tensors can be joined by a so-called dyadic product $R_{(\mathbf{i},\ \mathbf{j})} = T_{\mathbf{i}} \cdot U_{\mathbf{j}}$.

□

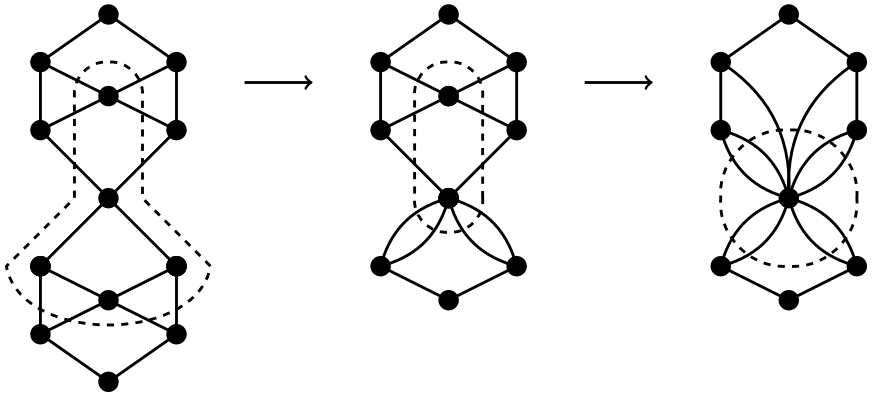
The following illustration shows examples of the diagrammatic notation. This notation is very useful when it comes down to understanding tensor equations, algorithms like the *ALS*, the *tensor train decomposition* and tensor subnetworks. It is also used in later chapters about error estimates e.g. when the maximal rank has to be estimated.



The above diagrams show some tensor networks using the graphical representation. From left to right: a scalar, a vector, a matrix, an order three tensor, a matrix vector equation $Ax = b$.



The above figure shows the decomposition of an order four tensor into the *tensor train format*. The dashed line separates the indices into two groups of indices which are flattened by using the matricisation.



The above figure shows a tensor network with a subnetwork (encircled with a dashed line) which is going to be contracted as far as possible. It is important to notice that all vertices where the connecting edges belong to the subnetwork are contracted in the first step. In the second step all vertices in the subnetwork which are not connected by edges belonging to the subnetwork are contracted.

Figure 1: The above drawings show some basic tensor networks and demonstrate the graphical representation of tensors explained before. Further they demonstrate the decomposition of tensors as well as the idea of subnetworks in a tensor network.

2.2 The Tensor Train Format

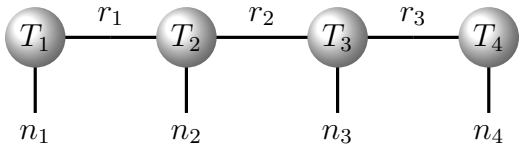
The storage of an order d tensor with homogeneous external indices n scales exponentially in the dimension $\mathcal{O}(n^d)$. This fact is called the *curse of dimensionality*. Tensor formats in general attempt to improve not only the storage scaling behaviour but also the operations connected to it such as scalar product, addition etc.
The tensor format which is going to be of special interest in this thesis is the *tensor train format* (TT format). The TT format has an interesting history as it was discovered several times by different fields of study. In physics, especially in the quantum many body community, the TT format is known as *Matrix Product States* (MPS) and has been used for numerical calculations since 1993 when White introduced the *DMRG* algorithm [4].
Mathematicians got in touch with the TT format when Oseledets started publishing about the format [5] and when Schneider, Holtz and Rohwedder introduced the so called *ALS* algorithm for optimisation problems in the tensor train format [6]. The following chapter is based on these works.
It starts with the definition of the *tensor train decomposition*.

Definition: An order d tensor T with external indices $i_j \in \{1, ..., n_j\}$ is given in the *tensor train format* if there exist tensors $T_1,\ T_d$ of order two and and tensors $T_3, ..., T_{d-1}$ of order three such that

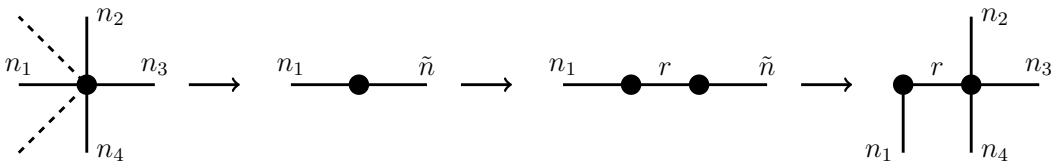
$$T_{\mathbf{i}} = \sum_{\alpha_1=1}^{r_1} \sum_{\alpha_2=1}^{r_2} ... \sum_{\alpha_{d-1}=1}^{r_{d-1}} T_{1,i_1,\alpha_1} T_{2,\alpha_1,i_2,\alpha_2} ... T_{d,\alpha_{d-1},i_d} \ .$$

The tuple $\mathbf{r} = (r_1,\ r_2, ..., r_{d-1})$ is called the *TT-rank* and the tensors T_i are the so-called *component tensors*.

It is important to notice that the indices α_0 and α_d can be added to the above definition. In that case they are one-dimensional. If these indices are mentioned or not, depends on the author and whether they are of interest or not. As they would complicate the following explanation of the external indices in the TT format and are not going to be of any use for this thesis, they will not be further mentioned.
The following shows a TT format of an order four tensor T with external indices $n_1, ..., n_4$. $T_1, ..., T_4$ describe the component tensors of T.



Under the assumption that the rank and the external indices are homogeneous the storage of such a tensor network scales as $\mathcal{O}(dnr^2)$. Hence it only scales linearly in the dimension. This is a huge improvement and breaks the curse of dimensionality. The TT-decomposition can be obtained by using the SVD for every of the d matricisations of the tensor (see figure 1 pic. 2). The tensor train decomposition is a tensor decomposition where the SVD is used partially on the tensor. Using the SVD of a matrix, its rank can be reduced by cutting of some of the singular values of the matrix. The rank of a matrix is directly connected to the rank of its tensorisation, at the respective index the tonsorisation was applied to. This connection is shown graphically in the following.



Hence transferring the procedure to construct a low-rank matrix approximation to the tensor train decomposition leads to a low rank approximation of a tensor. As the rank is directly connected to the bound dnr^2 of the storage of a tensor, a low-rank approximation has a smaller storage requirement.

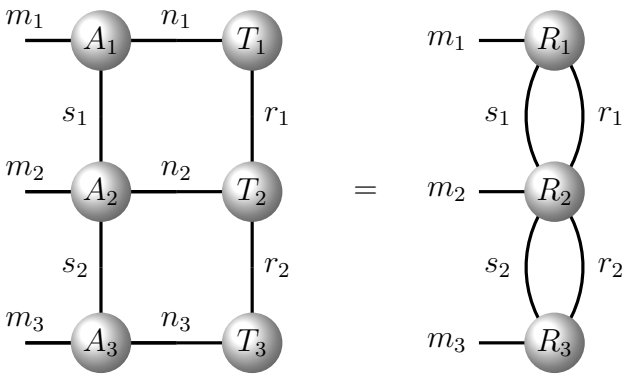
As the tensor train decomposition of a tensor enables us to improve its storage, an analogous format for tensor operators is required. The format that fulfils the rank and with it the storage reducing property is the *tensor train operator format*. As a natural generalisation of the tensor train format it will also be called tensor train format.

Definition: An order $2d$ tensor operator A with external indices $i_k \in \{1, ..., n_k\}$ and $j_k \in \{1, ..., m_k\}$ is given in the *tensor train format* if there exists tensors $A_1,\ A_{d-1}$ of order three and $A_2, ..., A_{d-1}$ of order four such that

$$A_{\mathbf{i},\mathbf{j}} = \sum_{\alpha_1=1}^{s_1} \sum_{\alpha_2=1}^{s_2} ... \sum_{\alpha_{d-1}=1}^{s_{d-1}} A_{1,i_1,j_1,\alpha_1} A_{2,\alpha_1,i_2,j_2,\alpha_2} ... A_{d,\alpha_{d-1},i_d,j_d} \ .$$

The tuple $\mathbf{s} = (s_1,\ s_2, ..., s_{d-1})$ is called the *TT-operator-rank* and the tensors A_i are the so-called *component tensors*.

The following shows an order three TT operator A acting on the order three TT tensor T. The result is an order three tensor R of rank $(r_0 \cdot s_0, ..., s_3 \cdot r_3)$.



2.3 Alternating Least Squares (ALS)

The *Alternating Least Squares* (ALS) algorithm aims to treat optimisation problems of a given functional by using the tensor train format. In this thesis the functional is $J(u) = ||b - Au||^2$. This general functional covers especially the approximation of eigenvectors and eigenvalues. They will be of special interest as measurable quantities of a physical system are eigenvalues. Here and in the following $||\cdot||$ denotes the euclidean norm and $\langle \cdot, \cdot \rangle$ the euclidean scalar product.
The following Lemma shows that the minimisation of $||b - Au||^2$ is equivalent to the minimisation of the functional $\langle u, Au \rangle - 2 \langle u, b \rangle$. Furthermore it reduces the original tensor network to a tensor network on which the *ALS* is easier to apply to.
In the following the existence of a solution of $Au = b$ is always assumed. The Lemma will be proven for the matrix case. This is a standard procedure as tensors can be matricised.

Lemma: Let A be a symmetric, positive definite matrix and b an arbitrary vector. Then the following holds:

$$Ax = b \quad \Leftrightarrow \quad x = \underset{u}{\operatorname{argmin}}(\langle u, Au \rangle - 2 \langle u, b \rangle)$$

Proof: We define $h(u) := \langle u, Au\rangle - 2\langle u, b\rangle$. Its critical points are given by the zeros of the derivative. Here

$$\begin{aligned} h'(u)v &= \langle u, Av\rangle + \langle v, Au\rangle - 2\langle v, b\rangle \\ &= 2\langle Au - b, v\rangle \ , \end{aligned}$$

hence $\nabla h(u) = 2(Au - b)$. This implies $Ax = b$. Showing that the Hessian is positive definite proves that x is a local solution

$$\frac{d^2}{du_i du_j} h(u) = \frac{d}{du_j}\left(2\sum_j u_j A_{j,i} - 2b_i\right) = 2A_{i,k} \ .$$

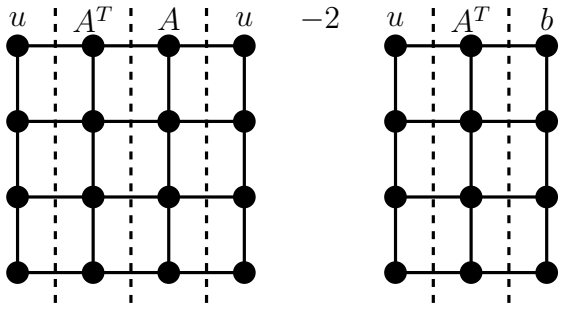
The convexity of h implies that this solution is also global.

□

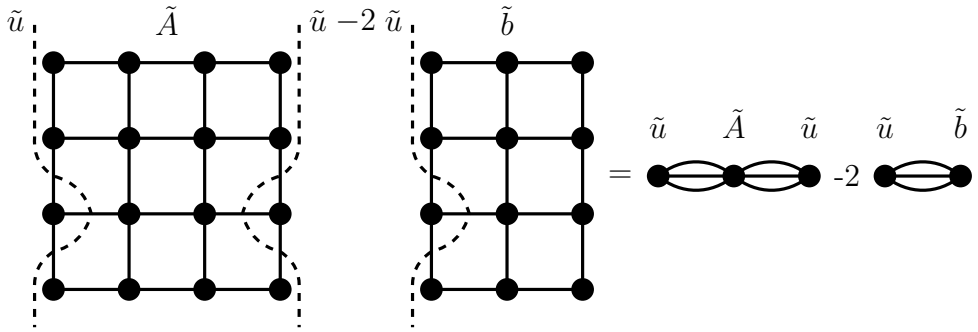
There are different ways of implementing the ALS. The implementation which is used in the following chapter exploits the following idea.
We are going to rewrite $\text{argmin}_u||b - Au||^2$ as follows

$$\begin{aligned} \text{argmin}_u||b - Au||^2 &= \text{argmin}_u \langle b - Au, b - Au\rangle \\ &= \text{argmin}_u(\langle b, b\rangle - 2\langle b, Au\rangle + \langle Au, Au\rangle) \\ &= \text{argmin}_u(\langle u, A^T Au\rangle - 2\langle u, A^T b\rangle) \ . \end{aligned}$$

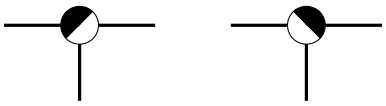
Visualising the system which is going to be minimised for a four dimensional problem the following graph is obtained.



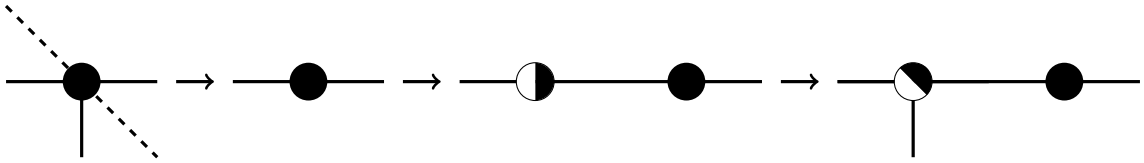
The ALS algorithm does not work on the entire system but on single components. The idea is to contract the system except for three components, one in the term $\langle u, A^T b\rangle$ and two in the term $\langle u, A^T Au\rangle$. The components which are not going to be contracted have the same index. This is visualised in the following.



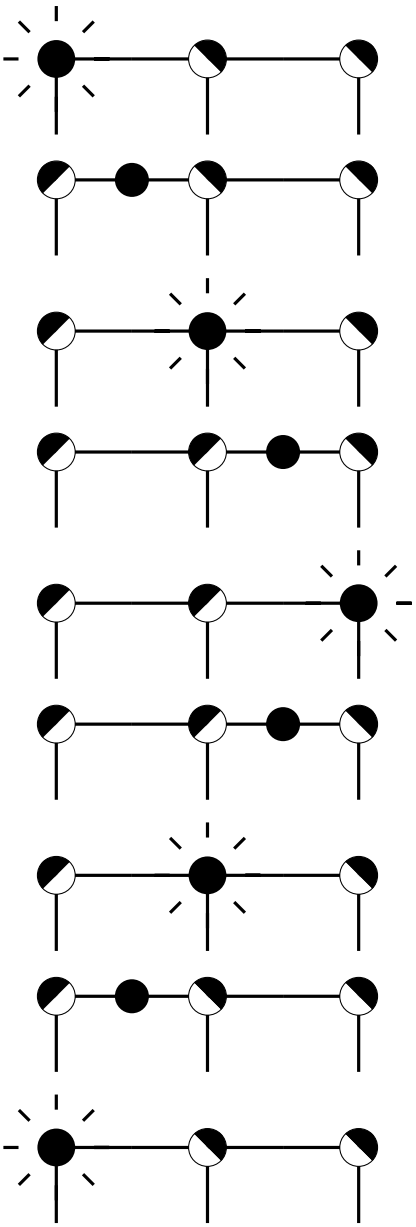
By construction $\tilde{A}$ is positive definite and symmetric. These quantities are preserved under matricisation. The expression $\langle \tilde{u}, \tilde{A}\tilde{u}\rangle - 2\langle \tilde{u}, \tilde{b}\rangle$ is going to be minimised with respect to $\tilde{u}$, which using the Lemma will solve $\tilde{A}\tilde{x} = \tilde{b}$. After the matricisation of $\tilde{A}$ this problem can be solved by any algorithm for linear problems. Considering a low rank approximation of the given problem this works quite fast as $\tilde{A}$ is an element of a $(r_i \cdot r_{i+1} \cdot n_i)^2$-dimensional vector space. It is important to notice that the reduced problem $\tilde{A}\tilde{x} = \tilde{b}$ is in general of much smaller dimension and therefore numerically more handy than the original problem. Solving this problem yields a vector $\tilde{x}$. In fact $\tilde{x}$ is a flattened order three tensor x where all indices are flattened together. Instead of rematricising the tensor and replacing the component tensor u_i in u directly by the tensor x it is first *left-orthonormalised*. In this context an order three tensor can be *left-* resp. *right-orthonormal* if the matricisation flattens the first resp. the last two indices yields an orthonormal matrix. In diagrammatic notation these object a represented by



We define that the external indices touching the vertex at the white part are flattened by the matricisation. Hence the left order three tensor is right-orthonormal and right order three tensor is left-orthonormal.
It is possible to obtain such a left-orthonormal tensor by matricising the tensor x where the first two indices are flattened and using a QR decomposition afterwards. This decomposition yields an order three tensors network containing two component tensors x_Q and x_R where x_Q is left-orthogonal. This procedure is depicted bellow.



Instead of replacing one component tensor u_i in u by one order three tensor x, it is replaced by such a tensor network.
To clarify these so-called *micro iterations* of the ALS it is helpful to only consider the changes in the tensor u. It is important to keep in mind that every micro iteration goes through the above explained procedure to create a tensor network which replaces the respective component tensor u_i of u. The assumption that all but the first component tensors u_i are right-orthonormal before an ALS *sweep* starts can be made w.l.o.g. The first *micro iteration* changes the first component tensor u_1 of u. A component tensor which is neither right-orthonormal nor left-orthogonal is called *core*. The *micro iteration* moves the core from the first component tensor u_1 to the second by replacing the first component tensor of u with the corresponding tensor network. The second *micro iteration* does the same procedure but on the second component tensor and so on. One *half sweep* ends if the core is moved to the last component tensor of u. The ALS starts the second half sweep by extracting the last component tensor of u and replacing it by the corresponding tensor network such that the last component tensor is not left- but right-orthonormal. Such a tensor network can be obtained by using a RQ decomposition instead of a QR decomposition of the matricisied tensor x. A sweep is finished if the core is moved to the first component tensor. A whole sweep is depicted in the following graph.



Initial situation: The first component tensor is the core.

First step of first micro iteration is complete: The first component tensor got replaced by the optimised tensor network

First micro iteration is complete: The component tensor x_R of the first tensor network got contracted with the second component tensor.

First step of second micro iteration is complete. The second component tensor got replaced by the optimised tensor network

Second micro iteration is complete such as the first half sweep. The core is now going to be moved to the left.

First step of first micro iteration of the second half sweep is complete: The last component tensor got replaced by the optimised tensor network.

First micro iteration of the second half sweep is complete.

First step of second micro iteration of the second half sweep is complete.

One sweep is complete. Inital situation for the next sweep.

In the following a pseudo code for the ALS is given.

Algorithm 1 . The ALS algorithm for the TT optimisation problem $Ax = b$

Require: $A,\ b,\ u,\ \epsilon_{req}$
Ensure: u such that $||Au - b|| \leq \epsilon_{req}$ holds.

```
ε ← ∞
while ε > ε_req do
  for i = 1, ..., d − 1 do
    Ã ← contraction of uA^T Au without u_i
    b̃ ← contraction of u^T A^T b without u_i
    solve Ãx̃ = b̃ for x̃
    (x_Q, x_r) ← QR-decomposition of matricisation of x̃
    u_i ← x_Q
    u_{i+1} ← x_R u_{i+1}
  end for
  for i = 1, ..., d − 1 do
    Ã ← contraction of uA^T Au without u_i
    b̃ ← contraction of u^T A^T b without u_i
    solve Ãx̃ = b̃ for x̃
    (x_r, x_Q) ← RQ-decomposition of matricisation of x̃
    u_i ← x_Q
    u_{i−1} ← x_R u_{i−1}
  end for
  ε ← ||Au − b||
end while
```

It is important to notice that the ALS requires a starting tensor u that fixes the rank. This is one of the downsides of the ALS because if the rank is chosen to low a solution will be impossible to find. If it is chosen to high the runtime gets to long. An algorithm that is rank-adaptive is the *modified ALS* (MALS).
The minimising quantity, the converging behaviour of the ALS and more information about the MALS can be seen more detailed in [6].

3 Spin Chains and the Hubbard Model

This chapter give brief introduction to the physical background of this thesis. It starts by explaining the *first quantisation* which comprises the basic ideas of quantum mechanical treatment of a system but only for one particle systems. This model is enlarged to describe a *many body system* in a quantum mechanical way. This is the *second quantisation*. The *second quantisation* of a *solid state physics system* leads to problems that can neither be solved analytically nor numerically. Therefore physical approximations like the *Hubbard model* are necessary to make these problems solvable. After a brief introduction to *solid state physics* and the involved solving problematic the *Hubbard model* is explained.

3.1 First Quantisation

The *first quantisation* describes the quantisation of a single particle moving through space-time. The starting point of this so called *canonical quantisation* process is *Hamiltonian mechanics*. It describes a system using a so-called *Hamiltonian* which represents the total energy of the system. It is given by

$$\mathcal{H} = T + V, \qquad \textit{with } T = \frac{p^2}{2m} \text{ and } V = V(q) \ . \tag{3.1}$$

Here T describes the kinetic energy and V the potential energy. In general $\mathcal{H}$ is a function of the generalised coordinates $p,\ q \in \mathbb{R}^3$. Often the notations $p(r,t)$ and $V(r,t)$ are used. In this case the equation above is dependant of $r,\ t$.
The basic idea of the first quantisation is to replace the energy and the momentum by operators. One quantum mechanical quantity is the *positional probability* given by

$$\mathbb{P}(r \in I) = \int_I |\psi(r,t)|^2 dr \ ,$$

where $\psi(r,t)$ is the *state function*. Hence the Hilbert space where the Hamiltonian is well-defined on has to be at least the space of the square-integrable functions $L^2(\mathbb{R}^4)$. The energy and the momentum are exchanged by the operators

$$\hat{E} = -\frac{\hbar}{i}\frac{\partial}{\partial t} \text{ and } \hat{p}_x = \frac{\hbar}{i}\frac{\partial}{\partial x} \ .$$

With (3.1) this leads to

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + V(x,t) \ ,$$

which is known as the *Hamilton operator*. Since functions of $L^2(\mathbb{R}^4)$ do not have to be differentiable the Hamilton operator is defined on $\mathcal{C}^2(\mathbb{R}^4) \cap L^2(\mathbb{R}^4)$. Here $\mathcal{C}^2(\mathbb{R}^4)$ is the space of all two times continuously differentiable functions on $\mathbb{R}^4$. As distributions like the δ-distribution play an important role in physics the above construct can and generally has to be done using *distributional differentiation*. As this construction only generalises the above idea and does not lead to more physical understanding of the first quantisation it is of no use for this thesis. Hence in the following no distributional argumentation is used.
Considering $\psi(r,t)$ as wave function of the quantum mechanical state depending on time t and space r

$$\hat{E}\Psi(r,t) = i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \left(\frac{-\hbar^2}{2m}\Delta + V(r,t)\right)\Psi(r,t) = \hat{H}\Psi(r,t) \tag{3.2}$$

holds. This is the *time-dependent Schrödinger equation*. This equation can be used in non-relativistic, single particle physics. As particles can have different charges their movement is affected by electromagnetic fields if they are present. Considering charged particles moving through electromagnetic fields, which are generally describes by *Maxwells equations* and the potentials A and ϕ, the equation (3.2) becomes

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \left(\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - qA\right)^2 + V + q\phi\right)\Psi(r,t)\ .$$

This is the *electromagnetic Schrödinger equation*. The eigenvalues and eigenfunctions are central quantities of any operator. In quantum mechanics measurable quantities are eigenvalues of the resp. operators. As the Hamilton operator describes the energy its eigenvalues are the measurable energies of a system. Therefore the eigenvalues of the Hamilton operator are of special interest for the description of a system. The resp. eigenvector is the state in which the system attains this energy.

3.2 Second Quantisation

To describe a quantum mechanical system a Hamiltonian and a Hilbert space on which the Hamiltonian is well-defined is needed. In the *second quantisation* these quantities are derived from the first quantisation.
It is important to notice that the following gives a basic introduction to the *second quantisation* but only to systems which are non-relativistic and consist of identical particles. The following procedure can be generalised to systems of distinguishable particles, see [7].

Considering a system of N identical particles (Fermions, Bosons etc.). Each of these particles has its own state space $\mathscr{H}_i$. The index indicates that a these states belong to one particular particle. As we are only going to consider systems in which the particles are indistinguishable there will only be one space defined using an index. The space of states for a system with a fixed number of N particles is given by

$$\mathscr{H}(N) = \bigotimes_{i=1}^{N} \mathscr{H}_1 \ .$$

The number in brackets behind the Hilbert space describes the actual number of particles which are involved in these states. It is important to notice that this only holds if the number of particles is fixed, which is rarely the case. Therefore the *Fock space* is defined as the space of states of a system with a varying number of particles. It is defined as the direct sum of the subspaces describing the states for fixed numbers of particles. Let the number of particles vary between zero and M. Then the *Fock space* is defined by

$$\mathscr{H} = \bigoplus_{N=0}^{M} \overline{S_\nu \ \mathscr{H}(N)} \ .$$

Here the operator S_ν is applied. This operator symmetrises or anti symmetrises a tensor and the overline represents the completion of the space. This linear operator is necessary because fermions and bosons obey different physical symmetry conditions. Further the degenerate $\mathscr{H}(0)$ is equal to $\mathbb{C}$.

The basic idea of constructing a Hamilton operator in terms of the second quantisation is to define operators that change the number of particles. They map $\mathscr{H}(N)$ to $\mathscr{H}(N \pm 1)$. These operators are so-called *creation* and *annihilation operators* and are denoted with $a^\dagger$ and a. They often depend on further physical quantities. Considering an atomic lattice the position where a particle is created respective annihilated is an important information just like the spin of the particle. In this case $a^\dagger_{i,\sigma}$ is the operator that creates a spin-σ particle on the lattice position i. The operator $a_{i,\sigma}$ is defined respectively.
The creation and annihilation operator are used to define the *number operator* $n = a^\dagger a$. This operator describes the number of particles in a *many body system*. This follows by the normalisation of $a^\dagger$ and a. As $a^\dagger$ and a depend on certain physical quantities, the number operator does as well. For each state $|\Psi\rangle \in \mathscr{H}$ of N particles there exist ladder operators $a^{r_1}_{i_1,\sigma_1}, ..., a^{r_q}_{i_q,\sigma_q}$ such that

$$|\Psi\rangle = a^{r_1}_{i_1,\sigma_1} \circ ... \circ a^{r_q}_{i_q,\sigma_q} |0_{\mathscr{H}}\rangle \ ,$$

where $r_m \in \{1, \dagger\}$ setting $a^1_{i,\sigma} := a_{i,\sigma}$. This representation gives rise to a reformulation of Hamiltonians using the creation, annihilation and number operators. This can be seen in detail in [7] [8].

3.3 Spin Chains

The spin chains considered in this thesis are one-dimensional, finite, loop free chains. Although these cases are very interesting in their application it is still unclear, how to apply the TT format to a closed spin chain or to a spin-lattice.
Such a spin chain simply consists of N sites, considering a spin-$\frac{1}{2}$ particle on each site. These particles can have either spin up $|\uparrow\rangle$ or spin down $|\downarrow\rangle$, hence any particle is in a linear state $\alpha|\uparrow\rangle + \beta|\downarrow\rangle$. This generates a local two-dimensional Hilbert space. Considering N fermions the state space, which is a Hilbert space, is given by

$$\mathscr{H} = \bigotimes_{i=1}^{N} \mathbb{C}^2 \ .$$

In the above argumentation only two spin states and their linear states were considered. A special case of a spin chain is the *Hubbard model*. Here states are a linear combination of four possible spin states. One lattice site can have no spin $|0\rangle$, one spin up $|\uparrow\rangle$, one spin down $|\downarrow\rangle$ or two spins up and down $|\uparrow\downarrow\rangle$. These states correspond to an *S-orbital* on one lattice site. It follows that any particle is in a linear state $\alpha|0\rangle + \beta|\uparrow\rangle + \gamma|\downarrow\rangle + \delta|\uparrow\downarrow\rangle$. For N lattice sites the corresponding Hilbert space is

$$\mathscr{H} = \bigotimes_{i=1}^{N} \mathbb{C}^4 \ .$$

It is important to notice that this is the Fock space that we obtain by using the procedure described in the section about second quantisation. As the elements of $\mathbb{C}^4$ are vectors it is important to notice that quantum mechanical states can be interpreted as tensors which were discussed before. Hence the atomic structures, atomic interactions of particles and possible physical optimisation problems are transferable to mathematical problems where the tensor structures like the TT format and the ALS are applicable.

An atomic lattice is an arrangement of particles whose movements are negligible. The simplest case of such a lattice is a one dimensional lattice, which is identified with a chain. The quantity $\mathbf{S}$ is the total spin angular momentum for all electrons belonging to one site. The tensor product of N of these spin states leads to a state in

$\mathscr{H}$. Why most of the fermionic systems can be described by a spin chain and how the Hamiltonians can be expressed for a general spin chain can been seen in [9].

3.4 The Hubbard Model

The *Hubbard model* is an approximate model used especially in *solid state physics*. It describes electrons on a fixed atomic lattice, considering only the *Coulomb interaction* of the electrons at the same lattice position.
As *solid state physics systems* are considered a brief introduction to *solid state physics* and why an approximation like the *Hubbard model* is useful is given in the following. After this brief introduction the *Hubbard model* is described. As a quantum mechanical model the Hubbard model is defined by a Hamilton operator and a Hilbert space of states where the operator is well-defined on. Therefore the model is going to be described using two paragraphs, one for the Hamiltonian and one for the Hilbert space.

3.4.1 Solid State Physics

Considering a *solid state body* its electrons are classified into two groups: the so-called *core electrons*, which are strongly bound to the atomic nucleus, and the *valence electrons*, which are not strongly bound and therefore moveable. The positive atomic nucleus and the negative core electrons are considered as one fixed object, the so-called *ion core* or in this context just *ion*. A solid state body consists of a three-dimensional crystal structure where the ions are arranged on a periodic atomic lattice. This lattice is considered to be fixed. As the ions consist of several electrons and the vaguely 10,000-times heavier atomic nucleus the valence electrons move much faster through the solid state body than the ions vibrate at their lattice position. Therefore the movement of the ions is negligible. The valence electrons move through the solid state body under the influence of the electronic potential $V(x)$ of all ions and the the Coulomb interaction. The Hamiltonian for N valence electrons of such a system is hence given by

$$\mathcal{H}_N = \sum_{i=1}^{N} \left(\frac{p_i^2}{2m} + V(x) \right) + \sum_{1 \leq i < j \leq N} U(x_i - x_j) \ ,$$

where x_i and p_i are the *position operator* and the *momentum operator* of the i-the valence electron. The expression $U(x_i - x_j)$ describes the electronic interaction between the valence electrons. This Hamiltonian leads to an analytically non-solvable eigenvalue problem. Therefore it is necessary to introduce certain models that provide

physical approximations. This means that physical constraints are used to make the system solvable. One of these models is the *Hubbard model*.
For a more detailed introduction into solid state physics see [10] [11] .

3.4.2 The Hamiltonian of the Hubbard Model

The construction of the Hamiltonian of the Hubbard model follows the steps of the second quantisation. First the creation and annihilation operators are defined. They are going to be identified with $a^\dagger_{i,\sigma}$ and $a_{i,\sigma}$. The index i describes the position in the lattice where a particle of spin σ is created resp. annihilated. As valence electrons, which are fermions, are considered, these operators have to satisfy the following commuting relations.

$$\begin{aligned} &[a_{i,\sigma}, a_{j,\tau}] = [a^\dagger_{i,\sigma}, a^\dagger_{j,\tau}] = 0 \qquad \text{and} \\ &[a_{i,\sigma}, a^\dagger_{j,\tau}] = \delta_{i,h}\delta_{\sigma,\tau} \ . \end{aligned}$$

Here the notation for the commutator $[A, B] = AB - BA$ was used. These operators define the *number operator* $n_{i,\sigma} = a^\dagger_{i,\sigma}\, a_{i,\sigma}$. This operator counts the electrons with spin σ at the lattice position i. The ladder operators and the number operator are used to define the Hamilton operator for the Hubbard model

$$\mathcal{H}_{Hub} = \sum_{i,j,\sigma\in\{\uparrow,\downarrow\}} t_{i,j} a^\dagger_{i,\sigma}\, a_{j,\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow} \ . \tag{3.3}$$

The complex numbers $t_{i,j}$ calculated by an integral formula which can be seen in [12] are the so called *hopping matrix elements*. As $\mathcal{H}_{Hub}$ is a hermitian operator these numbers have to fulfil the condition $t_{i,j} = \bar{t}_{j,i}$. The positive real number U is related to the coulomb interaction. For a more detailed discussion of these elements see [12] [11].
To illustrate the physical meaning of the single parts of the Hamiltonian it is useful to define the following operators

$$\begin{aligned} \mathcal{T} &= \sum_{i,j,\sigma} t_{i,j} a^\dagger_{i,\sigma\in\{\uparrow,\downarrow\}}\, a_{j,\sigma} \ , \\ \mathcal{V} &= U \sum_i n_{i,\uparrow} n_{i,\downarrow} \ . \end{aligned}$$

The operator $\mathcal{T}$ is called *tight binding Hamiltonian* and $\mathcal{V}$ is called the *on site interaction*.
Each summand of the tight binding operator describes the annihilation of an electron

with spin σ at the position j and the instantaneous creation of an electron with the same spin σ at the position i. This is called *quantum tunnelling* of an electron between two orbitals of different ions. A physical interpretation is therefore that $\mathcal{T}$ is the part of $\mathcal{H}_{Hub}$ which allows valence electrons to jump from one orbital to another independent of the ions and other valence electrons. This is called *delocalisation*. Hence it models the kinetic energy of non-interacting valence electrons.
As $t_{i,j} \to 0$ only the on site interaction remains in $\mathcal{H}_{Hub}$. The exact expression for $t_{i,j}$ given by [12] shows that $t_{i,j} \to 0$ is equivalent to considering an atomic lattice where the distance of the sites is infinite. This implies that there is no overlapping of the wave functions. Hence there can be no quantum tunnelling. Therefore $\mathcal{V}$ is the part of $\mathcal{H}_{Hub}$ that describes the localisation of electrons.
A more detailed characterisation of the Hamiltonian of the Hubbard model can be seen in [8], [10], [12], [13].

3.4.3 The Hilbert Space of the Hubbard Model

The Hubbard model describes the sites of the atomic lattice as one ion orbital that following the Pauli principle can contain the electron configurations on the i-th site of the atomic lattice:

- No electron. This is denoted by $|0\rangle$.
- One spin up electron.
  This is denoted by $|\uparrow\rangle$.
- One spin down electron.
  This is denoted by $|\downarrow\rangle$.
- Two electrons. One with spin up and the other one with spin down. This is denoted by $|\uparrow\ \downarrow\rangle$.

Assuming a lattice with L sites. According to the possible electron configurations such a lattice can contain between zero and $2L$ electrons. By N we denote the actual number of electrons in the lattice ($0 \leq N \leq 2L$) and $\mathbb{H}$ is the Hilbert space of one lattice site. As the lattice sites are indistinguishable only one one site Hilbert space is needed to construct the Fock space. As there are N sites on the lattice the Hilbert space of possible states is given by

$$\mathscr{H} = \bigotimes_{i=1}^{N} \mathbb{H} \ .$$

This Fock space is the same as constructed by the formalism given by the second quantisation. The advantages of this construction which is more oriented by the ideas of spin chains are that the tensor structure of the states follows directly and it does not use the one particle spaces. This further avoids the direct sum and leads directly to a representation of the state space by a tensor product.
There is more physical information about this Hilbert space covering *Bloch-bases*, *Wannier-states*, etc. which can be seen in [8], [10], [13].

4 Numerical Approach to the Hubbard Model

This chapter presents some numerical results of the ALS considering the functional $J(u) = ||\mathcal{H}_{Hub}x - x||$ by using the TT format. This functional is of interest because eigenvalues and eigenvectors are important physical quantities of a system (see chap. 3).
In the following we will restrict the Hamiltonian given by (3.3) to the special case where the hopping matrix elements and the positive number describing the coulomb interaction are set to one.
One of the standard algorithms to calculate eigenvectors is the *Rayleigh quotient iteration* [14], [15]. The *Rayleigh quotient iteration* algorithm is very similar to the *inverse iteration*, but replaces the estimated eigenvalue at the end of each iteration with the *Rayleigh quotient*. The algorithm starts by choosing a value μ_0 as an initial eigenvalue guess for the hermitian operator A. If this value is chosen close to the desired eigenvalue, then the algorithm approximates an eigenvector in that specific region. Also an initial vector b_0 must be chosen as initial eigenvector guess. Assuming that the approximations μ_i and b_i have already been computed the next approximation of the eigenvector b_{i+1} is then calculated by

$$b_{i+1} = \frac{(A - \mu_i Id)^{-1} b_i}{||(A - \mu_i Id)^{-1} b_i||} \ ,$$

where Id is the identity. The next approximation of the eigenvalue is set to the *Rayleigh quotient* of the current iteration,

$$\mu_i = \frac{\langle b_i \ , \ Ab_i \rangle}{||b_i||^2} \ .$$

As the Hamiltonian is an hermitian operator, this algorithm is applicable to $\mathcal{H}_{Hub}$. As tensor operators are used, the inverse in the iteration step turns out to be problematic. It is not clear how to calculate the inverse of a tensor but the ALS can be used to handle this problem. The ALS can be implemented to minimise the expression $||Au - b||$, which is equivalent to solving $Au = b$. As $(A - \mu_i Id)^{-1} b_i$ is the solution of $(A - \mu_i Id)x = b_i$ the ALS solves the problem and is therefore an essential part for the Rayleigh quotient iteration algorithm used in this thesis. In the following a pseudo code of the used version of the Rayleigh quotient iteration algorithm is given.

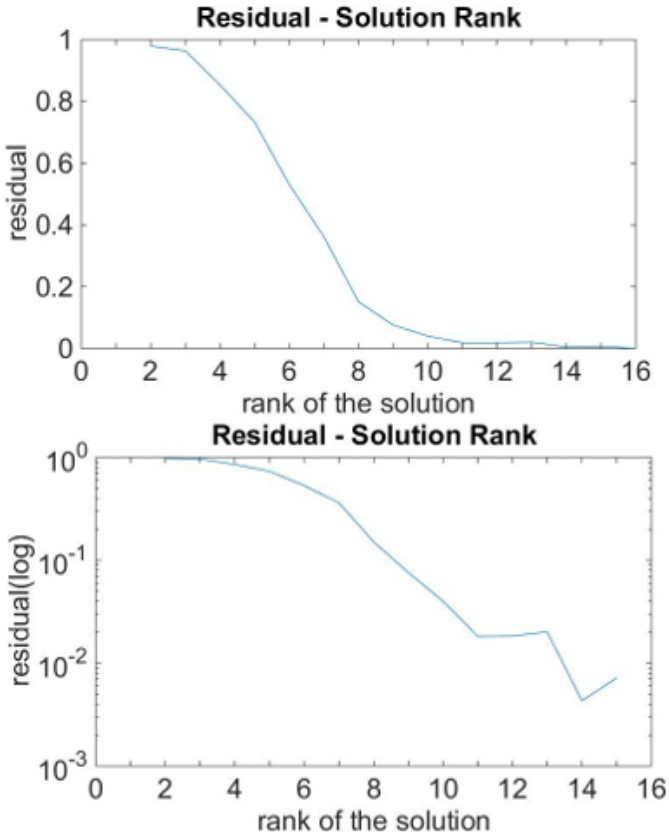
Algorithm 2 . The Rayleigh quotient iteration algorithm for operators in TT format

Require: $A,\ \epsilon_{req},\ r$

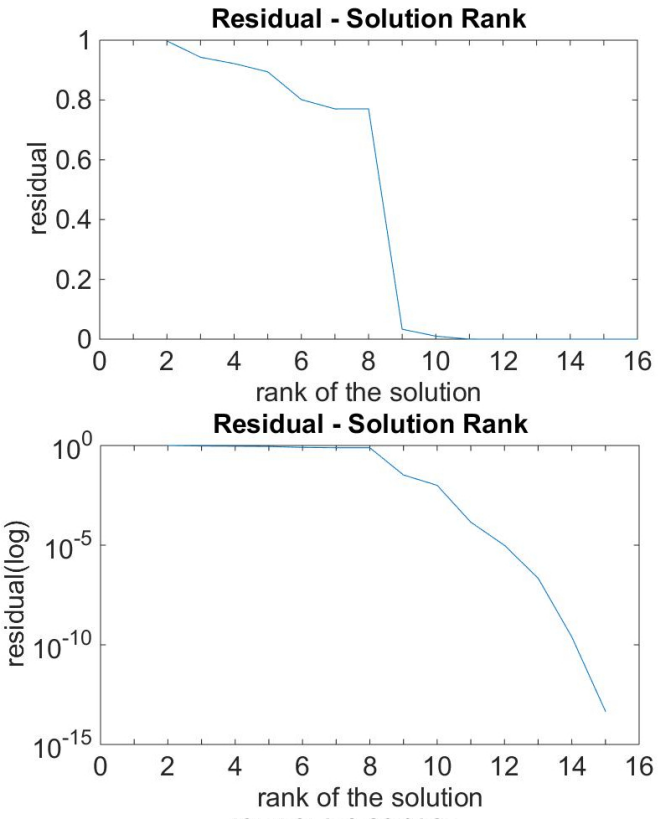
Ensure: Eigenvector u of A to the eigenvalue μ with iteration distance ϵ_{req}

$\epsilon = \infty$
$u \leftarrow$ random TT tensor of rank r
$\mu \leftarrow$ random number between 0 and 1
while $\epsilon > \epsilon_{req}$ **do**
  $u_h \leftarrow ALS(A - \mu Id, u, u, \epsilon_{ALS})$
  $u_h \leftarrow u_h \ / \ ||u_h||$
  $\mu \ \leftarrow u_h^T A u_h$
  $\epsilon \ \leftarrow ||u - u_h||$
  $u \ \leftarrow u_h$
end while

Knowing the eigenvector of a given operator like $\mathcal{H}_{Hub}$ allows an experimental analysis of the residual of the ALS while approximating this given eigenvector. Therefore we calculate the eigenvector of the considered Hamiltonian using the Rayleigh quotient iteration algorithm with a maximal rank r. This vector will be the solution b in the equation $\mathcal{H}_{Hub}u = b$. The ALS algorithm will minimise the functional $||\mathcal{H}_{Hub}u - b||$ where u will have different ranks. The following pictures shows the behaviour of the residual under varying approximation ranks of the solution.



In above figure a spin chain consisting out of five particles was considered. The Hamiltonian was constructed as in (3.3) and one eigenvector was calculated using the Rayleigh quotient algorithm. The solution rank variates from two to $(4, 16, 16, 4)$. In the diagram the maximal value of the rank characterises the abscissa.
In this section we focus on the behaviour of the residual in dependence of the solution rank. Therefore two diagrams are of interest. The data that compares the residual and the rank and the logarithmic plot of these data.
The above diagram shows the standard behaviour of the convergence of the ALS. The residual converges super linear until a certain rank is reached where it starts to converge linearly. This can be seen by looking at the logarithmic plots. This is a very important observation as it would be interesting to know on which condition this change of convergence behaviour depends.



The above diagrams show a different situation. The testing tensor is of order four. The rank varies in the same range as the tensor of order five but as the residual breaks in at approximately rank eight it is obvious that the eigenvector is not of maximal rank. Hence the solution does not always have to be of full rank. Therefore it would be interesting to know on which conditions rank of the solution depends.

These experimental results and the connected questions motivate this work. In following we will analyse the residual in dependence of the solution rank. Here the physical background will play a decisive role as it implies certain symmetries that simplify the considered operator.

5 Analysis of Residual Estimates

Eigenvalues and eigenvectors of the Hamiltonian are very important quantities of a physical system (see chap. 3). Hence a numerical approach to them is of special interest. Approximating the eigenvector comes down to solving the tensor equation $Ax = \lambda x$ which is a special case of $Ax = b$. The following chapter treats the error estimation of the residual of numerical approaches to the solution of such a tensor equation. The aim is to use numerical methods where the rank can be estimated in each iteration step. This leads to a bound of the residual by using error analysis for the numerical methods. In this thesis two numerical methods and one physical constrain are used. This section starts with a first error estimation using the *gradient method*. As the gradient method is one of the simplest methods and often not very efficient, the *Krylov subspace methods* are considered as well. Here the error estimation is based on a minimisation problem on the space of polynomials. To ensure the optimality of the estimation *Chebyshev polynomials* are used. To recall the necessary properties for the error estimation a brief introduction is given. As this error estimate for Krylov subspace methods can only be improved by considering special classes of operators, the physical background becomes of use. The Hubbard model and the nearest neighbour interaction approximation lead to an handy operator. Using combinatorial results, the error estimate for general operators given by the Krylov subspace methods can be improved.

5.1 Gradient Method Based Error Estimation

In this section the main interest is to prove a first bound for the residual in dependence of the tensor solutions' rank. Here the gradient method will be used. The chapter starts with a brief recapitulation of the gradient method. For a more detailed information see [16], [17].
The gradient method is an iterative method which aims to minimise any functional f.

Definition: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a given differentiable function. The gradient method is recursively defined by

$$x^{(k+1)} := x^{(k)} + t_k d^{(k)},$$

where $t_k > 0$ is the step size in the k-th iteration and $d^{(k)}$ is defined by
$d^{(k)} := -\nabla f(x^{(k)})$.

The ALS is implemented to minimise

$$f(x) = ||Ax - b||^2 \ . \tag{5.1}$$

Assume that A is a symmetric matrix, then

$$\nabla f(x) = 2A^T(Ax - b) = 2(A^2x - Ab) \ .$$

Oriented to this gradient method for the function f the iterative method for tensors

$$x^{(k+1)} := x^{(k)} - t_k 2(A^2x^{(k)} - Ab)$$

is defined. Here the following holds

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + t_k d^{(k)} = (Id - 2t_kA^2)x^{(k)} + 2t_kAb \\ &= (Id - 2t_kA^2)((Id - 2t_{k-1}A^2)x^{(k-1)} + 2t_{k-1}Ab) + 2t_kAb \\ &= ... \\ &= \left(\prod_{i=0}^{k}(Id - 2t_iA^2)\right)x^{(0)} + \sum_{i=0}^{k} 2t_i \left(\prod_{j=i+1}^{k}(Id - 2t_jA^2)\right)Ab \ . \end{aligned}$$

We define the map $P_A^{(k+1)} : \mathbb{T}(d, \mathbf{n}(A)) \to \mathbb{T}(d)$ as follows:

$$P_A^{(k+1)}(x) = \left(\prod_{i=0}^{k}(Id - 2t_iA^2)\right)x + \sum_{i=0}^{k} 2t_i \left(\prod_{j=i+1}^{k}(Id - 2t_jA^2)\right)Ab \ . \tag{5.2}$$

Hence $x^{(k+1)} = P_A^{(k+1)}(x^{(0)})$. We recall that for polynomials on $\mathbb{R}$

$$\prod_{i=0}^{n}(1 + a_ix) = \sum_{i=0}^{n+1}\alpha_i x^i, \qquad \textit{where } \alpha_i = \sum_{\{t_1,...,t_i\}\subseteq\{0,...,n\}} a_{t_1}...a_{t_i} \ .$$

This follows by induction. As the tensors A and Id commute the above result is applicable to $P_A^{(k)}$. Hence it follows that

$$\begin{aligned} \prod_{i=0}^{k-1}(Id - 2t_iA^2) &= \sum_{i=0}^{k}\alpha_iA^{2i} \qquad \textit{, where } \alpha_i = \sum_{\{j_1,...,j_i\}\subseteq\{0,...,k-1\}} (-2)^i t_{j_1}...t_{j_i} \\ \prod_{j=i+1}^{k-1}(Id - 2t_jA^2) &= \sum_{j=0}^{k-i-1}\beta_jA^{2j} \qquad \textit{,where } \beta_i = \sum_{\{j_1,...,j_i\}\subseteq\{i+1,...,k-1\}} (-2)^i t_{j_1}...t_{j_i} \ . \end{aligned}$$

Using this result in (5.2) yields

$$\begin{aligned} P_A^{(k)}(x^{(0)}) &= \left(\prod_{i=0}^{k-1}(Id - 2t_iA^2)\right)x^{(0)} + \sum_{i=0}^{k-1} 2t_i \left(\prod_{j=i+1}^{k-1}(Id - 2t_jA^2)\right) Ab \\ &= \sum_{i=0}^{k} \alpha_i A^{2i}x^{(0)} + \sum_{i=0}^{k-1} 2t_i \sum_{j=0}^{k-i-1} \beta_j A^{2j} Ab \ . \end{aligned} \tag{5.3}$$

We recall that for tensors operators the following holds:

$$\begin{aligned} \operatorname{rank}(A+B) &\leq \operatorname{rank}(A) + \operatorname{rank}(B) \\ \operatorname{rank}(A \cdot B) &\leq \operatorname{rank}(A) \cdot \operatorname{rank}(B) \ . \end{aligned}$$

Here $\cdot$ on the left-hand side denotes the contraction along all indices $\mathbf{n}(A)$ with all indices $\mathbf{m}(B)$. Further we recall for finite geometric series

$$\sum_{i=0}^{n} z^i = \frac{1-z^{n+1}}{1-z}, \qquad z \in \mathbb{C} \setminus \{1\} \ .$$

Defining $\operatorname{rank}(A) = r$, $\operatorname{rank}(x^{(0)}) = m$ and $\operatorname{rank}(b) = n$ the representation (5.3) of $P_A^{(k)}(x^{(0)})$ and the recalls above implies a sharp bound for the rank:

$$\begin{aligned} \operatorname{rank}(P_A^{(k)}(x^{(0)})) &\leq \sum_{i=0}^{k} r^{2i}m + \sum_{i=0}^{k-1}\sum_{j=0}^{k-i-1} r^{2j}rn \\ &= m\frac{1-r^{2k+2}}{1-r^2} + \sum_{i=0}^{k-1} nr\frac{1-r^{2k-2i}}{1-r^2} \\ &= m\frac{1-r^{2k+2}}{1-r^2} + \frac{nr}{1-r^2}\left(k - r^{2k}\frac{1-r^{-2k}}{1-r^{-2}}\right) \\ &= \frac{1}{1-r^2}\left(-mr^{2k+2} - nr^{2k+1}\frac{1-r^{-2k}}{1-r^{-2}} + nkr + m\right) \\ &=: \widetilde{r} \ . \end{aligned}$$

The approximation $x^{(k)}$ thus is a rank $\widetilde{r}$ approximation of the exact result $\widehat{x}$.
In general the residual of the gradient method for the functional (5.1) is given by

$$||\widehat{x} - x^{(k)}||_A \leq \alpha^k ||\widehat{x} - x^{(0)}||_A \ ,$$

where $||y||_A = \sqrt{\langle y, Ay\rangle}$ and α depends on A. As any tensor can be matricised this result is also applicable for the above iteration method for tensors.

The best approximation of rank $\tilde{r}$ will be written as $x_{best}^{(\tilde{r})}$ and

$$||\widehat{x} - x_{best}^{(\tilde{r})}|| \leq ||\widehat{x} - x^{(k)}||$$

For any rank $\tilde{r}$ approximation. Hence

$$||\widehat{x} - x_{best}^{(\tilde{r})}|| \leq \alpha^k ||\widehat{x} - x^{(0)}|| \ .$$

The asymptotic behaviour of $\tilde{r}$ in approximation steps k is $\tilde{r} \sim r^{2k}$ which means

$$\lim_{k \to \infty} \frac{\tilde{r}}{r^{2k}} = 1 \ .$$

Hence it follows that $2k \sim \log_r(\tilde{r})$, which is implies to

$$k \sim \frac{\log_r(\tilde{r})}{2} = \log_\alpha(\tilde{r}) \frac{1}{2 \log_\alpha(r)} \ .$$

In total $||\widehat{x} - x_{best}^{(\tilde{r})}|| \in \mathcal{O}\left(\tilde{r}^{\left(\frac{1}{2 \log_\alpha(r)}\right)}\right)$ holds for the residual of the gradient method.

5.2 Error Estimation for Krylov-Subspace Methods

The first bound was obtained using the gradient method. In this section the optimality property of the *Chebyshev polynomials* is going to be used to find an optimal bound for the so called *Krylov subspace methods*. Therefore this section starts with an introduction to the *Chebyshev polynomials* followed by an overview of the *Krylov-subspace-methods*. In the end a residual bound will be obtained using these tools.

5.2.1 Chebyshev Polynomials

A complete discussion of the *Chebyshev polynomials* is given in [18] [19]. In the following only those properties necessary for this work will be presented.
As this section is concerned with polynomials we start with the definitions of the relevant polynomial spaces.

Definition: The set of all polynomials of the degree n on the interval $[a, b]$ is denoted by

$$P_n[a,b] := \{p : [a,b] \to \mathbb{R},\ x \mapsto a_n x^n + a_{n-1}x^{n-1} + ... + a_0 x^0 | a_0, ..., a_n \in \mathbb{R}\} \ .$$

Further

$$\hat{P}_n[a,b] := \{p : [a,b] \to \mathbb{R},\ x \mapsto a_n x^n + a_{n-1}x^{n-1} + ... + a_0 x^0 | a_0, ..., a_n \in \mathbb{R},\ a_n = 1\}$$

is the set of all polynomials of the degree n and leading coefficient 1 on the interval $[a, b]$.

There are several ways of defining *Chebyshev polynomials*, we use the recursive version so that it is easy to see that the functions defined are indeed polynomials.

Definition: *Chebyshev polynomials* of the first kind are defined by the recurrence relation

$$\begin{aligned} T_0(t) &= 1 \\ T_1(t) &= t \\ T_{n+1}(t) &= 2tT_n(t) - T_{n-1}(t) \ . \end{aligned}$$

In the following it will be more convenient to use the following equivalent representation

$$T_n : [-1,1] \to \mathbb{R},\ t \mapsto \cos(n \arccos(t)) \ .$$

The equivalence is shown in [19]. To simplify the notation the Chebyshev polynomials of the first kind are going to be called Chebyshev polynomials.
The following Lemma shows some of their basic properties .

Lemma: For any Chebyshev polynomial T_n the following holds:

(i) $T_n(\cos(\theta)) = \cos(n\theta)$, for $\theta \in [0, \pi]$.

(ii) $\max_{t\in[-1,1]} |T_n(t)| = 1$

(iii) T_n has exactly $n+1$ extremal points $s_k^{(n)}$ on $[-1, 1]$, which are given by $s_k^{(n)} := \cos\left(\frac{k\pi}{n}\right)$. These points satisfy $T_n(s_k^{(n)}) = (-1)^k$ for $k = 0, ..., n$.

Proof:

(i) The first property follows from the invertibility of $\nu \mapsto \cos(\nu)$ on $[0, \pi]$. The inverse is given by $t \mapsto \arccos(t)$, thus $T_n(\cos(\theta)) = \cos(n\theta)$.

(ii) Furthermore $T_n(1) = \cos(n \cdot 0) = 1$ and using the boundedness of the cosine function we obtain $\max_{t\in[-1,1]} |T_n(t)| = 1$. This proves the second property.

(iii) To show the third property we define $s_k^{(n)} := \cos\left(\frac{k\pi}{n}\right)$ for $k = 0, ..., n$. Together with the first property we obtain

$$T_n(s_k^{(n)}) = T_n\left(\cos\left(\frac{k\pi}{n}\right)\right) = \cos\left(\frac{nk\pi}{n}\right) = \cos(k\pi) = (-1)^k \ .$$

Since $|\cos(x)| \leq 1$ for $x \in [-1, 1]$ all the critical points $s_k^{(n)}$ are in the interval $[-1, 1]$.

□

The following will be of utmost importance for the final optimality property of the Chebyshev polynomials used in this thesis.

Lemma: For any $n \in \mathbb{N}$ the polynomial

$$\tilde{p}(t) : [-1, 1] \to \mathbb{R},\ t \mapsto t^n - \frac{1}{2^{n-1}} T_n(t)$$

is the best approximation in $P_n[-1, 1]$ of the function $f_n : [-1, 1] \to \mathbb{R},\ t \mapsto t^n$ with respect to the supremum norm $||\cdot||_\infty$.

Before proving this Lemma we recall the so-called *Chebyshev alternation theorem*

Theorem: A polynomial p of degree $n \in \mathbb{N}$ is the best approximation of a function f on a compact set $K \subset \mathbb{R}$ if and only if n extremal points

$$x_0 < ... < x_n \in K$$

exists where

$$f(x_k) - p(x_k) = \sigma(-1)^k ||f - p||_\infty$$

for all $k = 0, ..., n+1$ and $\sigma = \pm 1$ fixed. These critical points are called an *alternant*.

The proof of this theorem can be seen in [20]. The alternation theorem will be the main tool for the proof of the above Lemma.

Proof: The $(n+1)$ critical points $s_k^{(n)} = \cos(\frac{k\pi}{n})$ of the $n-$th Chebyshev polynomial are an alternant for $d(t) := t^n - \tilde{p}(t) = \frac{1}{2^{n-1}} T_n(t)$. This follows as:

- $d(\cos\frac{k\pi}{n}) = \frac{1}{2^{n-1}}$, for k even.
- $d(\cos\frac{k\pi}{n}) = -\frac{1}{2^{n-1}}$, for k odd.
- $||d||_\infty = \max_{t\in[-1;1]} |\frac{1}{2^{n-1}} T_n(t)| = \frac{1}{2^{n-1}}$.

Using the alternation theorem and the fact that $d(s_k^{(n)}) = (-1)^k ||d||_\infty$ the polynomial $\tilde{p}$ of degree $n-1$ is a best approximation of f_n.

□

The following corollary shows that the polynomial $\frac{1}{2^{n-1}} T_n$ is the best approximation to the zero-function in $\hat{P}_n[-1, 1]$.

Corollary: For the n-th Chebyshev polynomial T_n

$$\left|\left| \frac{1}{2^{n-1}} T_n \right|\right|_\infty = \min_{p\in\hat{P}_n[-1,1]} [-1, 1]||p||_\infty \ .$$

Proof: As $\tilde{p}$ is the best approximation of f_n we obtain

$$\left|\left| \frac{1}{2^{n-1}} T_n \right|\right|_\infty = ||t^n - \tilde{p}||_\infty = \min_{p\in P_{n-1}[-1,1]} ||t^n - p||_\infty = \min_{p\in\hat{P}_n[-1,1]} ||p||_\infty \ .$$

□

Theorem: Let $\xi \notin [-1,1]$. Then

$$\min_{\substack{p \in P_n[-1,1], \\ p(\xi)=1}} ||p||_\infty = \left|\left| \frac{1}{T_n(\xi)} T_n \right|\right|_\infty$$

for the n-th Chebyshev polynomial.

Proof: In the following we set $t_n := \frac{1}{T_n(\xi)} T_n$. Assume there exists a polynomial $p \in P_n$ with $p(\xi) = 1$ and $||p||_\infty < ||t_n||_\infty$. The polynomial t_n differs from T_n only by a constant factor. By the first Lemma t_n attains its extreme values at $s_k^{(n)}$, $k = 0, ..., n$ on $[-1,1]$. The extreme values are alternating. Following the assumption $|p(s_k^{(n)})| < |t_n(s_k^{(n)})|$ for $k = 0, ..., n$. The polynomial $\omega := t_n - p$ has in total n changes of its sign, thus it has n zeros. Another zero of ω is at ξ because $\omega(\xi) = t_n(\xi) - p(\xi) = 1 - 1 = 0$ hence ω has $n+1$ pairwise disjoint zeros. Using the fact that ω is a polynomial whose degree is at most n it follows that ω is the zero-function. Hence $p = t_n$. This is in contradiction to the assumption that $||p||_\infty < ||t_n||_\infty$, which proves the claim.

□

The statement of the above theorem can be generalised to an arbitrary interval $[a,b]$, where the domain of the Chebyshev polynomials is stretched. This generalisation is described by the following corollary.

Corollary: Let $\xi \notin [a,b]$. We define $\phi : \mathbb{R} \to \mathbb{R}$ with $t \mapsto \frac{2t-a-b}{b-a}$ and

$$\tilde{T}_n : [a,b] \to \mathbb{R},\ x \mapsto T_n \circ \phi(x)\ .$$

Then

$$\min_{\substack{p \in P_n[a,b], \\ p(\xi)=1}} ||p||_\infty = \left|\left| \frac{1}{\widetilde{T}_n(\xi)} \widetilde{T}_n \right|\right|_\infty\ . \tag{5.4}$$

Proof: Let $[a,b]$ be an arbitrary interval and $\xi \notin [a,b]$. The function ϕ maps $[a,b]$ homeomorphically to $[-1,1]$. Indeed this function is an affine linear bijection. Therefore the operator

$$\Phi : P_n[-1,1] \to P_n[a,b],\ p \mapsto p \circ \phi$$

is well-defined and a bijection as well. This implies

$$\min_{\substack{p \in P_n[a,b],\\ p(\xi)=1}} ||p||_\infty = \min_{\substack{p \in P_n[-1,1],\\ (\Phi p)(\xi)=1}} ||\Phi p||_\infty = \min_{\substack{p \in P_n[-1,1],\\ p(\phi(\xi))=1}} ||p||_\infty = \left|\left|\frac{1}{T_n(\phi(\xi))} T_n\right|\right|_\infty$$
$$= \left|\left|\frac{1}{(\Phi T_n)(\xi)} \Phi T_n\right|\right|_\infty = \left|\left|\frac{1}{\widetilde{T}_n(\xi)} \widetilde{T}_n\right|\right|_\infty .$$

□

This is the most important property of the Chebyshev polynomials for this thesis. It states that the polynomial $\frac{1}{T_n(\phi(\xi))} T_n \circ \phi$ is the best approximation of the zero-function under the restriction that $p(\xi) = 1$ and $\xi \notin [a, b]$.

5.2.2 Krylov Subspace Methods

The properties shown above for the Chebyshev polynomials are now going to be very useful to estimate the error of the so-called *Krylov subspace methods*. A general abstract of the *Krylov subspace methods* can be seen in [16], [21].

Definition: *Krylov subspace methods* are methods used to solve a linear system $Ax = b$, where $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$. Defining $x^{(0)} \in \mathbb{R}^n$ as an arbitrary vector and $\mathcal{K}_k$, $\mathcal{L}_k$ as k-dimensional subspaces of $\mathbb{R}^n$. The so-called *Krylov subspace* $\mathcal{K}_k$ is defined by

$$\mathcal{K}_k := \mathcal{K}_k(A, r^{(0)}) = \text{Span}\left(r^{(0)}, Ar^{(0)}, ..., A^{m-1} r^{(0)}\right) ,$$

where $r^{(0)} := b - Ax^{(0)}$. The subspace $\mathcal{L}_k$ can be arbitrarily chosen. The Krylov subspace methods calculate an approximation $x^{(k)} \in x^{(0)} + \mathcal{K}_k$ to the solution $\hat{x}$ (if $\hat{x}$ exists) under the assumption that

$$(b - Ax^{(k)}) \perp \mathcal{L}_k .$$

Using the fact that $\mathcal{K}_k$ is defined by a linear hull, $r^{(k)} \in x^{(0)} + K_k$ can be written as

$$x^{(k)} = x^{(0)} + \sum_{i=0}^{k-1} \lambda_i A^i r^{(0)}, \quad \text{where } \lambda_i \in \mathbb{R} \text{ for } i \in \{0, ..., k-1\} .$$

Using the above definition we define analogously to the section above the map

$$P_A^{(k)}(x) := x - \sum_{i=0}^{k-1} \lambda_i A^i (b - Ax) \ .$$

Hence $x^{(k)} = P_A^{(k)}(x^{(0)})$. Analogous to the polynomial used in the *Gradient Method Based Error Estimation*, it can be shown that the bound of the rank

$$rank(P_A^{(k)}(x^{(0)})) \leq \sum_{i=0}^{k-1} r^i n =: \widetilde{r} \ ,$$

where the rank$(A) = r$ and the rank$(x^{(0)}) = n$ was assumed.

Lemma: Consider a Krylov subspace method to solve the equation $Ax = b$ with a regular matrix $A \in \mathbb{R}^{n \times n}$. Let $k \in \mathbb{N}$ be fixed, the columns of the matrix $V_k \in \mathbb{R}^{n \times k}$ be a basis of $\mathcal{K}_k$ and the columns of $W_k \in \mathbb{R}^{n \times k}$ be a basis of $\mathcal{L}_k$ such that $W_k^T A V_k \in \mathbb{R}^{k \times k}$ is regular. Then the Krylov subspace method yields the representation

$$x^{(k)} = x^{(0)} + V_k (W_k^T A V_k)^{-1} W_k^T r^{(0)} \ .$$

Proof: Since the columns of V_k are a basis of $\mathcal{K}_k$ there exists a unique $\alpha \in \mathbb{R}^k$ such that $x^{(k)} = x^{(0)} + V_k \alpha_k$. Using the orthogonality we obtain

$$W_k^T (b - A(x^{(0)} + V_k \alpha_k)) = 0 \ ,$$

which is yields

$$W_k^T A V_k \alpha_k = W_k^T (b - A x^{(0)}) \ .$$

As $W_k^T A V_k$ is regular this is equivalent to

$$\alpha_k = (W_k^T A V_k)^{-1} W_k^T r^{(0)} \ .$$

Hence the claim follows.

□

Using the above representation of the vector $x^{(k)}$ we obtain

$$r^{(k)} = b - A x^{(k)} = r^{(0)} - A V_k (W_k^T A V_k)^{-1} W_M^T r^{(0)} \ . \tag{5.5}$$

Before discussing the norm of $r^{(k)}$ a definition of a space of matrix polynomials is given.

Definition: We define

$$\widetilde{P}_k := \{p : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},\ A \mapsto a_k A^k + a_{k-1}A^{k-1} + ... + a_0 x A^0 | a_0, ..., a_k \in \mathbb{R},\ a_0 = 1\}$$

the set of all polynomials of the degree k and last coefficient 1 on the matrix space $\mathbb{R}^{n\times n}$. In this thesis $A^0 = Id$.

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be regular. Further let $\upsilon_1, ..., \upsilon_k \in \mathbb{R}^n$ and $\omega_1, ..., \omega_k \in \mathbb{R}^n$ be the basis vectors of $\mathcal{K}_k$ and $\mathcal{L}_k$ constructed by a Krylov subspace method. Defining the projection

$$P_k := Id - AV_k(W_k^T AV_k)^{-1}W_k^T$$

If $W_k^T AV_k \in \mathbb{R}^{k\times k}$, with $V_k = (\upsilon_1, ..., \upsilon_k) \in \mathbb{R}^{n\times k}$ and $W_k(\omega_1, ..., \omega_k) \in \mathbb{R}^{n\times k}$ is regular, then the estimate

$$||r^{(k)}|| \leq ||P_k|| \min_{p\in\widetilde{P}_k} ||p(A)r^{(0)}||$$

holds.

Proof: By the regularity of the matrix $W_k^T AV_k$ and the definition of P_k

$$P_k AV_k = AV_k - AV_k(W_k^T AV_k)^{-1}W_k^T AV_k = 0\ .$$

The representation (5.5) of the residual vector implies $r^{(k)} = P_k r^{(0)}$. Hence $r^{(k)} = P_k(r^{(0)} + AV_k\alpha)$ for any vector $\alpha \in \mathbb{R}^k$. The matrix V_k consists of basis vectors of the subspace $\mathcal{K}_k = \text{Span}\left(r^{(0)}, Ar^{(0)}, ..., A^{k-1}r^{(0)}\right)$ hence $AV_k\alpha \in A\mathcal{K}_k$. Thus $AV_k\alpha = \sum_{i=1}^k \lambda_i A^i r^{(0)}$ and therewith

$$r^{(k)} = P_k(r^{(0)} + AV_k\alpha) = P_k(r^{(0)} + \sum_{i=1}^k \lambda_i A^i r^{(0)}) = P_k \sum_{i=0}^k \lambda_i A^i r^{(0)}, \quad \text{where } \lambda_0 = 1\ .$$

It follows that $r^{(k)} = P_k p(A)r^{(0)}$ for any polynomial $p \in \widetilde{P}_k$. Hence

$$||r^{(k)}|| = \min_{p\in\widetilde{P}_k} ||P_k p(A)r^{(0)}|| \leq ||P_k|| \min_{p\in\widetilde{P}_k} ||p(A)r^{(0)}||\ .$$

□

In the following the space of polynomials of the degree n and last coefficient 1 on the interval $[a, b]$ will be needed. It is defined as follows.

Definition: We define

$$\widetilde{P}_k[a,b] := \{p : [a,b] \to \mathbb{R},\ x \mapsto a_k x^k + a_{k-1}x^{k-1} + ... + a_0 x^0 | a_0, ..., a_k \in \mathbb{R},\ a_0 = 1\}$$

the set of all polynomials of the degree k and last coefficient 1 on the interval $[a,b]$.

The above estimation of the norm of the residual vector implies the question, how small $\min_{p\in\widetilde{P}_k} ||p(A)||$ can get?
The standard procedure is to restrict the polynomials to some interval that contains the spectrum $\Lambda(A) = \{\lambda_1, ..., \lambda_\ell\}$ of A, where $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_\ell$ holds w.l.o.g. This follows by

$$\begin{aligned} ||p(A)||^2 &\overset{(1)}{=} \sup_{||x||=1} ||p(A)x||^2 = \sup_{||x||=1} \langle p(A)x, p(A)x\rangle \overset{(2)}{=} \sup_{||x||=1} \langle p^2(A)x, x\rangle = \rho(p^2(A)) \\ &= \max \Lambda(p^2(A)) \overset{(3)}{=} \max p^2(\Lambda(A)) = \max_{\lambda\in[\lambda_1,...,\lambda_\ell]} p^2(\lambda) \leq \max_{\lambda\in[\lambda_1,\lambda_\ell]} p^2(\lambda) \\ &\overset{(1)}{=} \left(\max_{\lambda\in[\lambda_1,\lambda_\ell]} p(\lambda)\right)^2 . \end{aligned}$$

In (1) we used that $t \mapsto t^2$ is monotonically increasing, which leads to the equality of the suprema. In (2) the symmetry of A was used. The equality (3) is shown in [22].The above equation is equivalent to $||p(A)|| \leq \left(\max_{\lambda\in[\lambda_1,\lambda_\ell]} p(\lambda)\right)^2$. Hence in the following we will consider the term

$$\min_{p\in\widetilde{P}_k[\lambda_1,\lambda_\ell]} ||p||_\infty . \tag{5.6}$$

As A is positive definite $\lambda_1 > 0$. This makes the property (5.4) applicable and hence leads to

$$\min_{p\in\widetilde{P}_k[\lambda_1,\lambda_\ell]} ||p||_\infty = \left|\left|\frac{1}{T_k(\phi(0))} T_k \circ \phi\right|\right|_\infty = \left|\left|\frac{T_k \circ \phi}{T_k\left(-\frac{\lambda_1+\lambda_\ell}{\lambda_\ell-\lambda_1}\right)}\right|\right|_\infty , \tag{5.7}$$

where $\phi : [\lambda_1, \lambda_\ell] \to [-1,1],\ t \mapsto \frac{2t-\lambda_1-\lambda_\ell}{\lambda_\ell-\lambda_1}$. The composition $T_k \circ \phi$ maps $[\lambda_1, \lambda_\ell]$ to $[-1,1]$ so the estimate

$$\left| T_k\left(\frac{2x-\lambda_1-\lambda_\ell}{\lambda_\ell-\lambda_\ell}\right)\right| \leq 1 , \tag{5.8}$$

where $\kappa(A)$ is the condition number of the matrix A with respect to the spectral norm. Furthermore

$$\frac{\lambda_1+\lambda_\ell}{\lambda_1-\lambda_\ell} = \frac{1+\frac{\lambda_\ell}{\lambda_1}}{1-\frac{\ell}{\lambda_1}} = \frac{1+\kappa(A)}{1-\kappa(A)} = \frac{1}{2}\left(\frac{1+\sqrt{\kappa(A)}}{1-\sqrt{\kappa(A)}} + \frac{1-\sqrt{\kappa(A)}}{1+\sqrt{\kappa(A)}}\right) . \tag{5.9}$$

For Chebyshev polynomials the property

$$T_k\left(\frac{1}{2}\left(x+\frac{1}{x}\right)\right)=\frac{1}{2}\left(x^k+\frac{1}{x^k}\right)\ , \tag{5.10}$$

follows by induction. We define $\gamma=\frac{1+\sqrt{\kappa(A)}}{1-\sqrt{\kappa(A)}}$. Inserting (5.8), (5.9) and (5.10) into (5.7) we obtain

$$\min_{p\in\widetilde{P}_k[\lambda_1,\lambda_\ell]}||p||_\infty\leq\left|\frac{1}{T_k\left(\frac{1}{2}\left(\gamma+\gamma^{-1}\right)\right)}\right|=2\mid\left(\gamma^k+\gamma^{-k}\ \right)^{-1}\mid=2\left|\left(\frac{\gamma^{2k}+1}{\gamma^k}\right)^{-1}\right|$$

$$=2\left|\frac{\gamma^k}{\gamma^{2k}+1}\right|\leq 2\ |\gamma^{-k}|=2\left|\left(\frac{1-\sqrt{\kappa(A)}}{1+\sqrt{\kappa(A)}}\right)^k\right|.$$

This implies the total estimate

$$||r^{(k)}||\leq||P_k||\ 2\left|\left(\frac{1-\sqrt{\kappa(A)}}{1+\sqrt{\kappa(A)}}\right)^k\right|\ ||r^{(0)}||=:\beta\ \alpha^k\ ||r^{(0)}||\ . \tag{5.11}$$

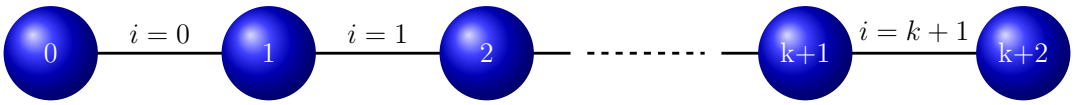
To clarify the consequence of this estimate for the rank estimate, we consider the asymptotic behaviour of the rank $\widetilde{r}\sim r^k$, which is equivalent to

$$k\sim\log_r(\widetilde{r})=\frac{\log_\alpha(\widetilde{r})}{\log_\alpha(r)}\ .$$

Thus $||r^{(k)}||\in\mathcal{O}\left(\widetilde{r}^{\left(\frac{1}{\log_\alpha(r)}\right)}\right)$. Due to the optimality of the Chebyshev-Polynomials and the fact that the above bound is sharp, this bound is optimal for Krylov subspace methods where no further assumptions are made to the operator considered.

5.3 Error Estimate Improvement by Using the Nearest Neighbour Interaction Approximation

The following section combines the physical approximation of a system using the nearest neighbour interaction with the results proven above. It is geared to [23].
In the following we are considering a physical system of $k+1$ particles, which are fixed between two *boundary-particles*. This system is pictured in the following.



As the energy of such a system is of interest the *nearest neighbour interaction approximation* is used as physical approximation of the system. It considers only the interaction between the nearest neighbours. The resulting operator is written in the following as a sum of operators each of which describes one interaction between two particles

$$A = \sum_{i=0}^{k+1} A_i \ .$$

In the following only systems where the Hamiltonian can be written as above are considered. The numerical methods used in this section are still the Krylov subspace methods. But physical constraints will be used to improve the given bound $\mathcal{O}\left(\widetilde{r}^{\left(\frac{1}{\log_\alpha(r)}\right)}\right)$.
According to (5.11) and (5.6) an estimate for the rank of A^n, where $n \in \mathbb{N}$, is of main interest. The operators A_i and $A_{i\pm1}$ do not commute. Hence an equation like

$$\sum_{\substack{j_0,\ldots,j_{k+1} \\ j_0+\ldots+j_{k+1}=n}} A_0^{j_0}\ldots A_{k+1}^{j_{k+1}} \neq (A_0+\ldots+A_{k+1})^n = A^n$$

is not possible. Therefore operators $\chi_{j_0,\ldots,j_{k+1}}$ are defined, which are given by a sum of products of $A_0,\ldots,A_{k+1}$, where for all $i \in \{0,\ldots,k+1\}$ the A_i appears exactly j_i times in the product. With these operators A^n can be written as

$$A^n = (A_0+\ldots+A_{k+1})^n = \sum_{\substack{j_0,\ldots,j_{k+1} \\ j_0+\ldots+j_{k+1}=n}} \chi_{j_0,\ldots,j_{k+1}} \ .$$

In the following the properties of these operators are discussed.

Lemma: For every multi-index $(j_0, ..., j_{k+1})$ with $j_0 + ... + j_{k+1} = n$ there exists an $w \in \{1, ..., k\}$ such that for any $s \in \mathbb{N}$ there exists a $d \in [w - s/2, w + s/2] \cap \mathbb{N}$ such that

$$j_d \leq \frac{n}{s} \ .$$

Proof: Assuming there exists a multi-index $(j_0, ..., j_{k+1})$ such that for all $w \in \{1, ..., k\}$ there exists a $s \in \mathbb{N}$ such that for all $d \in [w - s/2, w + s/2] \cap \mathbb{N}$ the inequality $j_d > \frac{n}{s}$ holds. Then

$$j_0 + ... + j_{k+1} = j_0 + j_{k+1} + \sum_{i=1}^{k} j_i \geq \sum_{d \in [w-s/2, w+s/2] \cap \mathbb{N}} j_d > s\frac{n}{s} = n \ .$$

This is a contradiction as by definition $n = j_0 + ... + j_{k+1}$.

□

Defining the operators $Q_{i,n,l}$, where $Q_{i,n,l}$ is the sum of those $\chi_{j_0,...,j_{k+1}}$, for which $j_i = l$ and $\sum_{m \neq i} j_m = n - l$ holds. The above Lemma states that for every multi-index there exists an w such that the multi-index contains a j_d, which is less or equal to n/s for any $s \in \mathbb{N}$. Hence $j_d \in \{0, 1, ..., \lfloor n/s \rfloor\}$ holds for one $d \in [w - s/2, w + s/2] \cap \mathbb{N}$. Demanding that $Q_{i,n,l}$ contains the sum of all $\chi_{j_0,...,j_{k+1}}$ implies that

$$\sum_{i \in [w-s/2, w+s/2] \cap \mathbb{N}} \sum_{l=0}^{\lfloor \frac{n}{s} \rfloor} Q_{i,n,l}$$

contains the sum

$$\sum_{\substack{j_0, ..., j_{k+1} \\ j_0 + ... + j_{k+1} = n}} \chi_{j_0,...,j_{k+1}}$$

which is A^n. Considering

$$Q_{i,n,l} = \sum_{\substack{j_i = l \\ \sum_{m \neq i} j_m = n - l}} \chi_{j_0,...,j_{k+1}} \ .$$

In the following properties of the operator $Q_{i,n,l}$ are discussed. To make a clear statement about the rank of these operators the following definitions

$$P_n^{(i)} := (L + A_i + R)^n, \qquad \text{with} \quad L := \sum_{m<i} A_m \text{ and } R := \sum_{m>i} A_m$$

are used. Considering only the terms of $P_n^{(i)}$ where A_i appears exactly l times, we obtain $Q_{i,n,l}$. Thanks to the commutativity of L and R these terms are of the shape

$$L^{\lambda_0} R^{\rho_0} A_i \; L^{\lambda_1} R^{\rho_1} A_i \; ... \; L^{\lambda_{l-1}} R^{\rho_{l-1}} A_i \; L^{\lambda_l} R^{\rho_l}, \qquad \text{with} \quad \sum_{j=0}^{l} (\lambda_j + \rho_j) = n - l \; .$$

Before discussing more properties of $Q_{i,n,l}$, a Lemma is introduced.

Lemma: Let $g, \; h \in \mathbb{N}$. Then

$$\sum_{\alpha=1}^{g} \binom{g - \alpha + h - 1}{h - 1} = \binom{g + h - 1}{h}$$

holds.

Proof: The proof of this claim follows by induction.

Basis $g = 1$: It holds

$$\sum_{\alpha=1}^{1} \binom{1 - \alpha + h - 1}{h - 1} = \binom{h - 1}{h - 1} = 1 = \binom{h}{h} = \binom{1 + h - 1}{h} \; .$$

Inductive step $g \to g + 1$: It holds

$$\begin{aligned} \sum_{\alpha=1}^{g+1} \binom{g + 1 - \alpha + h - 1}{h - 1} = \sum_{\alpha=0}^{g} \binom{g - \alpha + h - 1}{h - 1} &= \binom{g + h - 1}{h - 1} + \sum_{\alpha=1}^{g} \binom{g - \alpha + h - 1}{h - 1} \\ &\underset{\text{(I.V.)}}{=} \binom{g + h - 1}{h - 1} + \binom{g + h - 1}{h} = \binom{g + h}{h} \; , \end{aligned}$$

hence it follows the statement.

□

This Lemma is fundamental for the following theorem which states the number of summands in the operator $Q_{i,n,l}$. The residual bound estimated in this section is based on combinatorial arguments. Hence the number of summands is fundamental for these arguments.

Theorem: $Q_{i,n,l}$ has $t := \binom{n+l+1}{2l+1}$ summands.

Proof: In the regarded terms of $P_n^{(i)}$ there are l operators A_i. These are always found in a product with L and R. In total there are $2l+2$ products of the form $L^{\lambda_d}R^{\rho_d}$ and with these respectively $l+1$ powers λ_d and ρ_d, where $\sum_{j=0}^{l}(\lambda_j+\rho_j) = n-l$ holds. Hence there are $2l+2$ summands, which sum to the result $n-l$.

In the following the claim, that for h summands, which sum up to g, there are exactly $\binom{g+h-1}{h-1}$ possibilities, where we assume that g and all of the h summands have the same sign. The proof of this claim follows by induction.

Basis $h = 1$: For $h = 1$ $\binom{g+1-1}{1-1} = \binom{g}{0} = 1$ holds. This is obviously true because for one summand there is only one option to sum up to g, which is that the summand itself is g.

Inductive step $h \to h+1$: For $h+1$ there are first of all all optional summands to sum g up out of h summands. This is true because the last summand can be fixed to zero. Thanks to the inductivity there are already

$$\binom{g+h-1}{h-1}$$

options to sum g out of $h+1$ summands. The remaining options demand that the last summand is unequal to zero. Hence, the last summand value $\alpha = 1, ..., g$. The first h summands have to sum up to $g-\alpha$, where $\alpha = 1, ..., g$ holds. Again the induktivity and the above Lemma yield

$$\sum_{\alpha=1}^{g}\binom{g-\alpha+h-1}{h-1} = \binom{g+h-1}{h}.$$

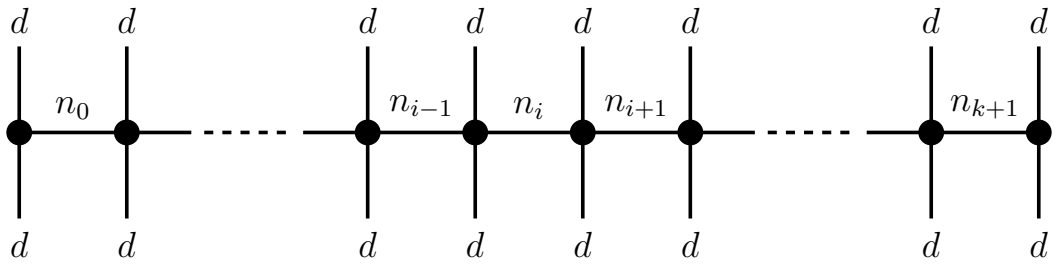
In total this leads to

$$\binom{g+h-1}{h-1} + \binom{g+h-1}{h} = \binom{g+h}{h} = \binom{g+(h+1)-1}{(h+1)-1}$$

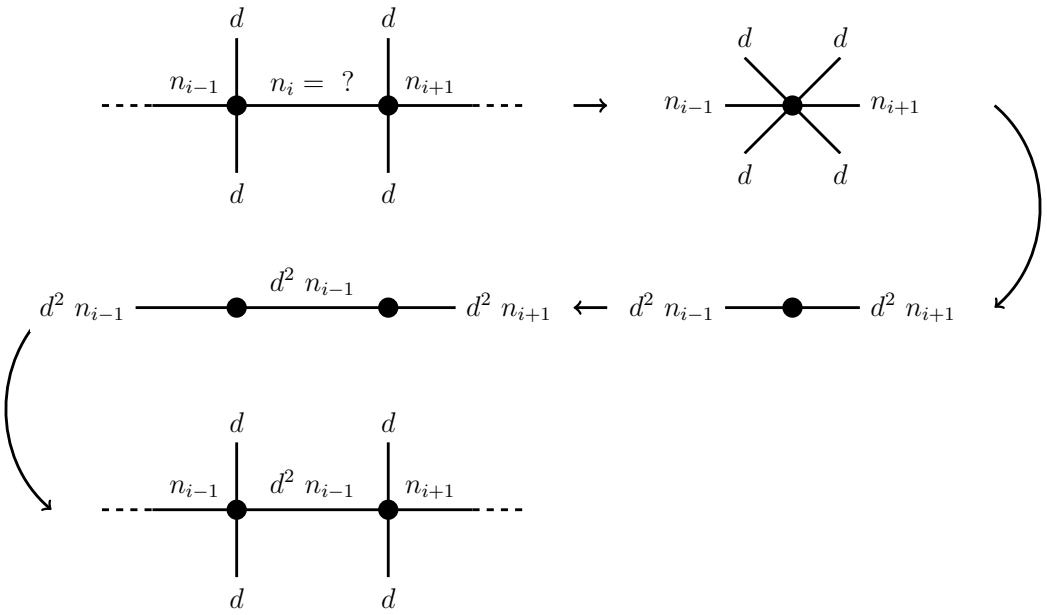
possible options to sum g out of $h+1$ summands. This proves the statement. With $g = n - l$ and $h = 2l + 2$ the claim follows for the operator $Q_{i,n,l}$.

□

Considering an arbitrary tensor operator of order $(k+2)$ with $2(k+2)$ external indices.



The operators A_i are elements of this class of Tensors. The rank n_i can be estimated as follows. Knowing $n_{i-1},\ n_{i+1}$ and d the rank n_i can be bound by a combination of the rank and the dimension to the right or to the left of n_i. Contracting along n_i we obtain a tensor of order six. The matricisation of this tensor is a matrix of the size $d^2\ n_{i-1} \times d^2\ n_{i+1}$. Using an SVD decomposition a tensor network is re-established. The rematricisation of this network leads to the original structure where n_i is bound by $d^2 n_{i-1}$. This process is illustrated in the following.



As A_i does not change the state of the j-th particle, where $j \neq i$ the rank $n_j = 1$. Hence $\text{rank}(A_i) = d^2$. In $Q_{i,n,l}$ there are l operators A_i which are multiplicatively connected. Furthermore the nearest neighbour interaction yields that these are the only operators that influence the rank at the position i. Hence it follows that $\text{rank}(Q_{i,n,l})$ at the position i is less or equal to d^{2l}. Such an estimation can be improved by a low rank approximation of A_i. W.l.o.g. we assume that $n_i = R^2$ with $R \leq d$ holds. Hence the rank of $Q_{i,n,l}$ at the position i is bound by R^{2l}.

Knowing the rank at the position i of $Q_{i,n,l}$ an estimate of the rank in the middle $k/2$ of the tensor network $Q_{i,n,l}$ is possible. Estimating this rank yields an upper bound to the global rank because the highest rank is in the middle of the tensor network. The procedure is the same as described above but this time we want to estimate the rank n_{i+1}, where w.l.o.g. we assume that $i < k/2$. This is bound by $d^2\ R^{2l}$. We now continue this iteration until the index $k/2$ is reached. Hence it is bound by

$$n_{k/2} \leq (d^2)^{|i-k/2|} R^{2l} \leq d^s R^{2l} \ .$$

In the last step we have used that $i \in [w - s/2, w + s/2] \cap \mathbb{N}$. Hence the maximal distance between two position indices is s. The operator $Q_{i,n,l}$ is composed of $\binom{n+l+1}{2l+1}$

operators, where each rank can be bound by $R^{2l}d^k$. It follows

$$\text{Rank}(Q_{i,n,l}) \leq \binom{n+l+1}{2l+1} R^{2l} d^k \ .$$

To get a rank estimation without the binomial coefficient the following Lemma is needed.

Lemma: For $n,\ k \in \mathbb{N}$ with $n, k \geq 0$ and $n \geq k$ it holds

$$\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k \ .$$

Proof: The e-function is defined by a power-series. Here the estimation

$$e^k = \sum_{l=0}^{\infty} \frac{k^l}{l!} \geq \frac{k^k}{k!} \Leftrightarrow k! \geq \left(\frac{k}{e}\right)^k$$

holds. Hence it follows

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)...(n-k+1)}{k!} \leq \frac{n^k e^k}{k^k} = \left(\frac{ne}{k}\right)^k \ .$$

□

These tools defined and proven here lead to a better rank estimation of A^n, which then leads to a better bound for the residual using the Krylov subspace methods. Considering

$$\begin{aligned}
\text{rank}(A^n) &\leq \text{rank}\left(\sum_{i\in[w-s/2,w+s/2]\cap\mathbb{N}} \sum_{l=0}^{\lfloor \frac{n}{s} \rfloor} Q_{i,n,l}\right) \leq \sum_{i\in[w-s/2,w+s/2]\cap\mathbb{N}} \sum_{l=0}^{\lfloor \frac{n}{s} \rfloor} \binom{n+l+1}{2l+1} R^{2l} d^s \\
&\leq \sum_{i\in[w-s/2,w+s/2]\cap\mathbb{N}} \sum_{l=0}^{\lfloor \frac{n}{s} \rfloor} \left(e\frac{n+l+1}{2l+1}\right)^{2l+1} R^{2l} d^s \leq \sum_{l=0}^{\lfloor \frac{n}{s} \rfloor} (e(n+2))^{2l+1} R^{2l} d^s \\
&\leq \left(\frac{n}{s}+1\right) s(e(n+2))^{2n/k+1} R^{2n/k} d^s \\
&\leq (n+s)(e(n+2))^{2n/s+1} R^{2n/s} d^s \ .
\end{aligned}$$

As s is arbitrary we assume $s = \sqrt{n}$. Furthermore $s \leq k$ holds. This implies

$$\operatorname{rank}(A^n) \leq (n + \sqrt{n})(e(n+2))^{2\sqrt{n}+1} R^{2\sqrt{n}} d^{\sqrt{n}} \ .$$

Using the representation of x_m given by the Krylov subspace method

$$\begin{aligned} \operatorname{rank}(x_m) &= \operatorname{rank}\left(x_0 + \sum_{i=0}^{m-1} \lambda_i A^i b - \sum_{i=1}^{m} \lambda_{i-1} A^i x_0 \right) \\ &\sim \operatorname{rank}(x_0) + \operatorname{rank}(A^{m-1} b) + \operatorname{rank}(A^m x_0) \end{aligned}$$

holds. In the above equation the last summand is dominating. Assuming that $\operatorname{rank}(x_0) = r$ holds the equation

$$\operatorname{rank}(x_m) \lesssim (m + \sqrt{m})(e(m+2))^{2\sqrt{m}+1} R^{2\sqrt{m}} d^{\sqrt{m}} r =: \widetilde{r}$$

holds. Considering the asymptotic behaviour of this rank in m

$$\begin{aligned} (m + \sqrt{m})(e(m+2))^{2\sqrt{m}+1} R^{2\sqrt{m}} d^{\sqrt{m}} &\leq 2m(2em^3)^{2\sqrt{m}+1} \leq 2(2em^4)^{2\sqrt{m}+1} \\ &\leq 2m^{12\sqrt{m}+6} \sim m^{12\sqrt{m}+6} = e^{\ln(m)(12\sqrt{m}+6)} \end{aligned}$$

holds. In this equation $d,\ R \leq m$ was assumed. Further $\ln(m) \in \mathcal{O}(m^\epsilon)$ is used, where $\epsilon > 0$ holds. This yields

$$e^{\ln(m)(12\sqrt{m}+6)} \in \mathcal{O}\left(e^{12m^{\epsilon+1/2}+m^\epsilon}\right) \subseteq \mathcal{O}\left(e^{13m^{\epsilon+1/2}}\right) \ .$$

Using this result

$$\widetilde{r} \sim e^{13m^{\epsilon+1/2}} \Leftrightarrow \frac{\ln(\widetilde{r})}{13} \sim m^{\epsilon+1/2} \Leftrightarrow m \sim \left(\frac{\ln(\widetilde{r})}{13}\right)^{\frac{2}{2\epsilon+1}}$$

holds. Using the analysis for the residual given by the Krylov subspace methods, the above calculations imply

$$||r_m|| \leq \beta \alpha^m ||r_0|| \in \mathcal{O}\left(\alpha^{\left(\frac{\ln(\widetilde{r})}{13}\right)^{\frac{2}{2\epsilon+1}}}\right) \ .$$

This is a improvement to the bound found for the Krylov subspace methods without assuming any physical approximations to the system.

6 Approximability

The following chapter is going to discuss the question of approximability of a physical system. It is geared to [24]. To discuss this question properly more physical quantities are needed. In many body systems a so-called *density operator* is an important quantity. It describes the probability of a system to be in a certain state as well as any possible state given by a many body system. This state can be *pure* or *mixed*. It is important to notice that a *mixed state* is not a *quantum mechanical superposition of states*.

Definition: Let $\mathscr{H}$ be an infinite state space with an orthonormal basis $(|\psi_i\rangle)_i$ of *pure states*. Then the *density operator* ρ can be defined as

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \ .$$

The quantity p_i describes the probability of the system to be in the *pure state* $|\psi_i\rangle$. The $(p_i)_i$ are a probability distribution. This implies the normalisation $1 = \sum_i p_i$. W.l.o.g. we assume that the p_i are monotonically decreasing. When it comes down to approximation of a physical system the *Renyi entropy* is an important quantity.

Definition: Let $\mathscr{H}$ be an infinite state space with an orthonormal basis $(|\psi_i\rangle)_i$ of *pure states* and the density operator ρ given by $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. Then the *Renyi entropy* is given by

$$S^\alpha(\rho) = \frac{1}{1-\alpha} \log(Tr(\rho^\alpha)) = \frac{1}{1-\alpha} \log\left(\sum_i p_i^\alpha\right) \ .$$

Here $Tr(\cdot)$ describes the trace operator.

In the following the operator norm

$$||\cdot|| := \frac{1}{\sqrt{n}} ||\cdot||_F$$

is used. Here $||\cdot||_F$ describes the *Frobeniusnorm*.
The following Lemma leads to an upper bound of the residual in dependants of the singular values of the matricisations of the tensor.

Lemma 1: Let $|\psi\rangle$ be a state describing a system of N particles where each particle has a d-dimensional state space. Then there exists a rank r TT format of this state $|\psi_r\rangle$ such that

$$|||\psi\rangle - |\psi_r\rangle|| \leq \frac{1}{\sqrt{d^N}} \sum_{\alpha=1}^{N} \epsilon_\alpha(r) \ .$$

Proof: The given state $|\psi\rangle$ is an order N tensor with homogeneous external indices of dimension d. By using partially the SVD on the matricisations of this tensor we get a TT format of $|\psi\rangle$ with maximal rank. We recall that the SVD of a matrix is the best low-rank approximation if we cut of some singular values. This property will be used by the projections P_i. This map projects the given tensor in TT format to a rank $\mathbf{r}_i = r$ TT tensor. Using the optimality property this leads to a tensor which is the optimal $\mathbf{r}$-approximation if no further projections are used. This procedure can be done for all $i = 1, ..., N$ to obtain a TT approximation of the given state $|\psi\rangle$. It is important to notice, that

$$||(Id - P_1)T|| \leq \frac{1}{\sqrt{d^N}} \epsilon_1(r), \qquad \text{where } \epsilon_j(r) = \left(\sum_{i=r+1}^{\infty} \sigma_{j,i} \right)^{\frac{1}{2}}$$

holds. The quantity $\sigma_{j,i}$ is the i-th singular value of the matricisation with respect to the index j. These facts yield

$$\begin{aligned}
||T - B|| &= ||T - P_1P_2P_3...T|| = ||T - P_1T + P_1T - P_1P_2P_3...T|| \\
&\leq ||T - P_1T|| + ||P_1T - P_1P_2P_3...T|| \\
&\leq ||(Id - P_1)T|| + ||P_1||\ ||T - P_2P_3...T|| \\
&= \frac{1}{\sqrt{d^N}} \left(\sum_{i=r+1}^{N_1} \sigma_{1,i} \right)^{\frac{1}{2}} + \underbrace{||P_1||}_{\leq 1}\ ||T - P_2T + P_2T - P_2P_3...T|| \\
&\leq \frac{1}{\sqrt{d^N}} \epsilon_1(r) + ||T - P_2T|| + ||P_2T - P_2P_3...T|| \\
&\leq \frac{1}{\sqrt{d^N}} \epsilon_1(r) + \frac{1}{\sqrt{d^N}} \epsilon_2(r) + ||P_2||||T - P_3...T|| \\
&\leq \frac{1}{\sqrt{d^N}} \left(\sum_{\alpha=1}^{N} \epsilon_\alpha(r) \right) \ .
\end{aligned}$$

□

This Lemma shows that for any state, thus also for any tensor, there exists a rank r TT approximation such that the residual is smaller then the sum of all singular values which had to be cut-off to reduce the respective rank from r_i to r.
To combine the Renyi entropy and the above given residual we are going to need the following Lemma.

Lemma 2: Let ρ be a given density operator. For any $\alpha \in (0,1)$ the equation

$$\log(\epsilon(r)) \leq \frac{1}{1-\alpha}\left(S^{\alpha}(\rho) - \log\left(\frac{r}{1-\alpha}\right)\right)$$

holds.

Proof: We are first going to define the quantity

$$p = \sum_{i=r+1}^{\infty} p_i$$

which can be identified with $\epsilon(r)$ as the p_i are connected with the singular values of ρ. We assume that p and therefore $r \in \mathbb{N}$ is a given quantity. We are first going to show that a probability distribution minimising the Renyi entropy has the following shape

$$\begin{aligned} p_1 &= 1 - p - (r-1)h \\ h &= p_2 = p_3 = ... = p_{r+p/h} \\ 0 &= p_{r+p/h+1}, p_{r+p/h+2}, ..., p_{\infty} \ . \end{aligned} \tag{6.1}$$

We are first going to minimise the Renyi entropy by changing the density only on the first r values. This follows by induction. Let therefore $r = 2$ be fixed. Then we know that the first the values of the density can be written as $p_3 = c,\ p_2 = c + b$ and $p_1 = c + a$. The first step is to show the following inequality

$$(c+a+b)^{\alpha} + c^{\alpha} \leq (c+a)^{\alpha} + (c+b)^{\alpha} \ . \tag{6.2}$$

Assuming $\alpha = 1$ or $\alpha = 0$ the claim follows directly. Let $\alpha \in (0,1)$. To show the above equation the function

$$h : \mathbb{R}^2 \to \mathbb{R},\ (a,b) \mapsto (c+a+b)^{\alpha} + c^{\alpha} - (c+a)^{\alpha} - (c+b)^{\alpha}$$

is defined. Obviously $h(0,0) = 0$, which means that in (6.2) the equality holds. Showing that all directional derivative out of $[0,1]^2$ in direction of $v \in \mathbb{R}^2_+$ are negative proves the claim. Considering therefore the gradient of h. Here

$$\nabla h(a,b) = \begin{bmatrix} \alpha((c+a+b)^{\alpha-1} - (c+a)^{\alpha-1}) \\ \alpha((c+a+b)^{\alpha-1} - (c+b)^{\alpha-1}) \end{bmatrix} \ .$$

Let $v \in \mathbb{R}^2_+$ and $(a,b) \in [0,1]^2$. Then

$$\langle \nabla h(a,b)\ ,\ v\rangle = \overbrace{\alpha}^{\geq 0}(\underbrace{((c+a+b)^{\alpha-1}-(c+a)^{\alpha-1})}_{\leq 0}\overbrace{v_1}^{\geq 0}+\underbrace{((c+a+b)^{\alpha-1}-(c+b)^{\alpha-1})}_{\leq 0}\overbrace{v_2}^{\geq 0})$$
$$\leq 0\ .$$

This proves (6.2).
Let $r \in \mathbb{N}$. The aim is still to create a distribution that minimises the Renyi entropy by only changing the first r values. As we have shown before, we can already assume that the density is of the shape $p_1 = c+a,\ p_2 = ... = p_r = c+b$ and $P_{r+1} + c$. As in the case of $r = 2$ the inequality

$$(c+a+(r-1)b)^\alpha + (r-1)c^\alpha \leq (c+a)^\alpha + (r-1)(c+b)^\alpha \tag{6.3}$$

proves the claim. Proving this inequality follows the same idea as proving (6.2). Defining

$$h : \mathbb{R}^2 \to \mathbb{R},\ (a,b) \mapsto (c+a+(r-1)b)^\alpha + c^\alpha - (c+a)^\alpha - (r-1)(c+b)^\alpha\ .$$

The Gradient is given by

$$\nabla h(a,b) = \begin{bmatrix} \alpha((c+a+(r-1)b)^{\alpha-1} - (c+a)^{\alpha-1}) \\ \alpha(r-1)((c+a+(r-1)b)^{\alpha-1} - (c+b)^{\alpha-1}) \end{bmatrix}\ .$$

With similar arguments to the proof of equation (6.2) the statement

$$\langle \nabla h(a,b)\ ,\ v\rangle \leq 0$$

follows. $h(0,0) = 0$ holds hence it follows the claim. This proves the equation (6.3). Considering only the first r values a probability density that minimises the Renyi entropy is of the shape

$$p_1 = 1 - p - (r-1)p_r, \qquad p_2 = ... = p_r\ .$$

As we first considered the first r values we are now going to consider the distribution values from p_{r+1}. Considering an arbitrary, monotonically decreasing probability distribution $(p_i)_{i\in\mathbb{N}}$. Further we are considering the distribution

$$\begin{aligned} \tilde{p}_i &= p_i, \qquad i \in \{1,...,r-1\} \\ \tilde{p}_i &= p_r, \qquad i \in \left\{r, ..., \frac{1-\sum_{i=1}^r p_i}{p_r}\right\}\ . \end{aligned}$$

We are going to show that

$$\sum_{i=1}^{\infty} p_i^\alpha \geq \sum_{i=1}^{\infty} \tilde{p}_i^\alpha \tag{6.4}$$

holds. As the first r terms are the same we are considering

$$\begin{aligned}\sum_{i=r+1}^{\infty} p_i^\alpha - \sum_{i=r+1}^{\infty} \tilde{p}_i^\alpha &= \sum_{i=r+1}^{\infty} p_i^\alpha - (1-\sum_{i=1}^{r} p_i)\frac{p_r^\alpha}{p_r} = \sum_{i=r+1}^{\infty} p_i^\alpha - \underbrace{\frac{p_r^\alpha}{p_r}}_{\leq 1} + \sum_{i=1}^{r} p_i \\ &\geq \sum_{i=r+1}^{\infty} p_i^\alpha + \sum_{i=1}^{r} p_i \frac{p_r^\alpha}{p_r} - 1 \geq \sum_{i=r+1}^{\infty} p_i + \sum_{i=1}^{r} p_i \frac{p_r}{p_r} - 1 \\ &= \sum_{i=1}^{\infty} p_i - 1 = 0 \ .\end{aligned}$$

Hence (6.4) follows.
Combining these results the Renyi entropy minimising probability distribution has to be of the shape (6.1). The Renyi entropy of such a probability density can be lower bound. Considering

$$\begin{aligned}\sum_i p_i^\alpha &= (1-p-(r-1)h)^\alpha + \left(r-1+\frac{p}{h}\right)h^\alpha \\ &= (1-p-(r-1)h)^\alpha - h^\alpha + rh^\alpha + ph^{\alpha-1} \ .\end{aligned}$$

Taking a closer look to $(1-p-(r-1)h)^\alpha - h^\alpha$ yields

$$(1-p-(r-1)h)^\alpha - h^\alpha = (1-\underbrace{(p+rh)}_{\substack{\leq 1 \\ (*)}} - h)^\alpha - h^\alpha \geq 0 \ .$$

In $(*)$ $p_1 \geq h$ was used which leads to $p + rh \leq p + p_1 + (r-1)h = 1$. Hence

$$\sum_i p_i^\alpha \geq rh^\alpha + ph^{\alpha-1} \ .$$

Defining $f(h) := rh^\alpha + ph^{\alpha-1}$ we can find an h independent lower bound for the Renyi entropy. The extremal points of f are given by

$$0 \stackrel{!}{=} f'(h) = \alpha h^{\alpha-1}\left(rh + p - \frac{p}{\alpha}\right) \Leftrightarrow h_{1,2} = \frac{p-\alpha p}{\alpha r},\ 0 \ .$$

As $h_2 = 0$ implies that $p_1 = 1$ and $p_i = 0$ for all $i \geq 2$ only the point h_1 is of interest. Verifying that f attains in h_1 a minimum follows by

$$\begin{aligned} f''(h)\Big|_{h_1} &= (\alpha-1)h^{\alpha-3}(\alpha r h + (\alpha-2)p)\Big|_{h_1} \\ &= (\alpha-1)\left(\frac{p-\alpha p}{\alpha r}\right)^{\alpha-3}\left(\alpha r\frac{p-\alpha p}{\alpha r} + (\alpha-2)p\right) \\ &= \underbrace{(\alpha-1)}_{\leq 0}\underbrace{\left(\frac{p-\alpha p}{\alpha r}\right)^{\alpha-3}}_{\geq 0}\underbrace{p\,(1-2)}_{\leq 0} \geq 0 \ . \end{aligned}$$

This implies directly

$$\begin{aligned} f(h) \geq f(h_1) &= r\left(\frac{p-\alpha p}{\alpha r}\right)^{\alpha} + p\left(\frac{p-\alpha p}{\alpha r}\right)^{\alpha-1} \\ &= r^{\alpha-1}p^{\alpha}\left(\frac{(1-\alpha)^{\alpha}}{\alpha^{\alpha}} + \frac{\alpha^{1-\alpha}}{(1-\alpha)^{1-\alpha}}\right) \\ &= \frac{r^{\alpha-1}p^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\left((1-\alpha)^{\alpha}(1-\alpha)^{1-\alpha} + \alpha^{1-\alpha}\alpha^{\alpha}\right) \\ &= \frac{r^{\alpha-1}p^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}} \ . \end{aligned}$$

Remembering the definition of the Renyi entropy this implies

$$S^{\alpha}(\rho) = \frac{1}{1-\alpha}\log\left(\sum_i p_i^{\alpha}\right) \geq \frac{1}{1-\alpha}\log\left(\frac{r^{\alpha-1}p^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right) \ .$$

Using this lower bound for the Renyi entropy it is possible to find an upper bound for the quantity p which will lead to the claim of the Lemma.

$$\begin{aligned} S^{\alpha}(\rho) &\geq \frac{1}{1-\alpha}\log\left(\frac{r^{\alpha-1}p^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right) \\ &= \frac{\alpha}{1-\alpha}\left(\frac{(1-\alpha)}{\alpha}\log\left(\frac{r}{1-\alpha}\right) - \log(\alpha) + \log(p)\right) \\ \Leftrightarrow \frac{(1-\alpha)}{\alpha}S^{\alpha}(\rho) - \frac{(1-\alpha)}{\alpha}\log\left(\frac{r}{1-\alpha}\right) &\geq -\log(\alpha) + \log(p) \geq \log(p) \ . \end{aligned}$$

This is equivalent to

$$p \leq \exp\left(\frac{(1-\alpha)}{\alpha}\left(S^{\alpha}(\rho) - \log\left(\frac{r}{1-\alpha}\right)\right)\right) \ .$$

□

To emphasise the consequences of this Lemma we are considering a spin chain of the length $N = 2L$ with a given Hamiltonian $\mathcal{H}$. $|\psi_{ex}\rangle$ denotes the ground state of this system. The entropy of such a system scales as

$$S^{\alpha}(\rho_L) \simeq \frac{c+\overline{c}}{12}\left(1+\frac{1}{\alpha}\right) ln(L) + \frac{\kappa}{N}$$

[25]. The aim is to ensure for any given $\epsilon_0 > 0$ the existence of a TT approximation of the ground state such that

$$|| \ |\psi_{ex}\rangle - |\psi_r\rangle||^2 \leq \frac{\epsilon_0}{L} \ . \tag{6.5}$$

It is important to notice that ϵ_0 is independent of L and the rank r is not fixed. It is obvious that a maximal rank TT approximation always satisfies (6.5) but the scaling of r is of special interest as a maximal rank approximation is not computable for large systems. We are therefore going to prove a statement about the minimal necessary rank such that (6.5) holds.
We define r_L as the minimal rank such that (6.5) holds for a spin chain of length N. Further we define $\epsilon(r) := \max_{i\in\{1,\dots,N\}} \epsilon_i(r)$. Using the Lemma 1 and 2

$$\begin{aligned}
|| \ |\psi_{ex}\rangle - |\psi_r\rangle||^2 &\leq \frac{N}{\sqrt{d^{N/2}}}\epsilon(r) \leq \frac{N}{\sqrt{d^{N/2}}} exp\left(\frac{1-\alpha}{\alpha}\left(S^{\alpha}(\rho_L) - ln\left(\frac{r_L}{1-\alpha}\right)\right)\right) \\
&\simeq \frac{N}{\sqrt{d^{N/2}}} exp\left(\frac{1-\alpha}{\alpha}\left(\frac{c+\overline{c}}{12}\left(1+\frac{1}{\alpha}\right) ln(L) + \frac{\kappa}{N} - ln\left(\frac{r_L}{1-\alpha}\right)\right)\right) \\
&= \frac{N}{\sqrt{d^{N/2}}} L^{\frac{1-\alpha}{\alpha}\ \frac{c+\overline{c}}{12}\left(1+\frac{1}{\alpha}\right)} e^{\frac{1-\alpha}{\alpha}\frac{\kappa}{N}} \left(\frac{1-\alpha}{r_L}\right)^{\frac{1-\alpha}{\alpha}} =: \tau \ .
\end{aligned}$$

To ensure that (6.5) holds we assume $\tau \leq \frac{\epsilon_0}{L}$. This implies

$$\begin{aligned}
&\frac{N}{\sqrt{d^{N/2}}} L^{\frac{1-\alpha}{\alpha}\ \frac{c+\overline{c}}{12}\left(1+\frac{1}{\alpha}\right)} e^{\frac{1-\alpha}{\alpha}\frac{\kappa}{N}} \left(\frac{1-\alpha}{r_L}\right)^{\frac{1-\alpha}{\alpha}} \leq \frac{\epsilon_0}{L} \\
&\Leftrightarrow \left(\frac{N}{\sqrt{d^{N/2}}}\right)^{\frac{\alpha}{1-\alpha}} L^{\frac{c+\overline{c}}{12}\left(1+\frac{1}{\alpha}\right)} e^{\frac{\kappa}{N}} \left(\frac{1-\alpha}{r_L}\right) \leq \left(\frac{\epsilon_0}{L}\right)^{\frac{\alpha}{1-\alpha}} \\
\Leftrightarrow &\left(\frac{N}{\sqrt{d^{N/2}}}\right)^{\frac{\alpha}{1-\alpha}} L^{\frac{c+\overline{c}}{12}\left(1+\frac{1}{\alpha}\right)} e^{\frac{\kappa}{N}} (1-\alpha) \left(\frac{\epsilon_0}{L}\right)^{\frac{1-\alpha}{\alpha}} \leq r_L \ .
\end{aligned}$$

This shows that r_L only has to scale polynomial in L such that (6.5) holds. It is important to notice that this result was shown under the assumption that the Renyi entropy scales logarithmically in L.

7 Conclusion

The idea of this work was to combine physics and mathematics to simulate physical systems, approximate eigenvalues of such systems and analyse the reliability of the numerical methods utilised for their computation. For the realisation of this aim the tensor train format played an essential role.
The numerical simulations were programmed in `MATLAB`. Further operations like contraction, permutation etc. on the tensor space were programmed independently. Furthermore the ALS using the functional $j(u) = ||Au - b||$ such as the Rayleigh quotient iteration method was implemented independently. For the tensor train format the toolbox by Oseledets was used.
Instead of using the exact Hamiltonian we use a simplified model making simulations feasible. Here the Hubbard model seemed an adequate choice. It is not only an academic model as there are systems that fit exactly into the framework established by it and at the same time it is not too complex to simulate. Realisations of such Hamiltonians were computed with `MATLAB`.
The combination of the numerical simulation of the Hubbard model and the implemented ALS were used to make quantitative statements about the behaviour of the residual in dependence of the solution rank. These results were described in chap. 3. The obtained quantitative statements were a motivation to describe the behaviour of the residual analytically. Therefore a numerical method, where the rank could be estimated in each iteration step was needed. In this thesis the Krylov subspace methods were used. Before considering a general Krylov subspace method the gradient method was considered to find a first bound of the residual.
In contrast to the gradient method the error analysis of the Krylov subspace methods transformed the residual estimate to an optimisation problem on the space of polynomials. Here Chebyshev polynomials were of utmost importance because they are the best approximation on the polynomial space with fixed grade of the zero function with respect to the standard norm. This led to a bound which under no further assumptions to the system is optimal for Krylov subspace methods. It is given by

$$||r^{(k)}|| \in \mathcal{O}\left(\widetilde{r}^{\left(\frac{1}{\log_\alpha(r)}\right)}\right) \ ,$$

where $r^{(k)}$ is the residual after k iteration steps, r is the solution rank and α is a constant which depends on the used operator.
As already mentioned this bound is optimal if no more constraints are made to the considered system. Inspired by the Hubbard model, spin chains and the therewith often combined nearest neighbour interaction constraint were used. These constraints

made combinatoric approaches possible such that a better estimation of A^n was possible. This estimation was used to improve the residual bound of Krylov subspace methods for the special case of a constrained physical system. This bound is given by

$$||r^{(k)}|| \in \mathcal{O}\left(\alpha^{\left(\frac{\ln(\widetilde{r})}{13}\right)^{\frac{2}{2\epsilon+1}}}\right) .$$

Furthermore the approximability of such a physical system was discussed. It was connected to the scaling of the Renyi entropy of a system. We proved that a quantum many body systems of solid state physics is approximable if the Renyi entropy scales logarithmically.

Comparing the calculated bounds to the numerical results of the ALS it is obvious that they are not optimal. This can be explained by the fact, that the bound is only optimal for Krylov subspace methods. As ALS is no Krylov subspace method the bounds calculated above do not hold. Hence a further question is which numerical methods lead to a smaller bound for the error estimate. Moreover the calculations in this thesis were often reduced to the maximal element of the solution rank. An exact treatment of the rank could also lead to a more precise and maybe lower bound.
As the basic structure the tensor train format was used. This is only one option for a low rank approximation of a tensor. It is important to notice that there are other tensor formats which allow such an approximation as well. The effects of physical constraints to different tensor formats could be considered as well and may lead to a more profound understanding of how physical constraints can be combined with mathematical structures.
In terms of physical constraints this work only considered loop free, non closed spin chains consisting of identical particles, thus one-dimensional simple objects. In physics especially in molecular dynamics the objects have not to be that simple. The chains might be closed, consisting of different particles with different properties or might be of higher dimensions e.g spin lattices. Further questions would be how the properties and proofs presented in this work can be generalised for such systems.
Spin chains can be treated in many different ways, the Hubbard model is only one of them. Hence different physical constraints could lead to lower bounds.

References

[1] W. Heisenberg. *Die Kopenhager Deutung der Quantentheorie*. Stuttgart. Battenberg, 1963.

[2] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer Verlag, 2012.

[3] L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM Journal on Matrix Analysis and Applications*, 2010.

[4] S. R. White. Density-matrix algorithms for quantum renormalization groups. *Physical Review*, 1993.

[5] I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 2011.

[6] S. Holtz, T. Rohwedder, and R. Schneider. The alternating linear scheme for tensor optimization in the tensor train format. *SIAM Journal on Scientific Computing*, 2012.

[7] W. Nolting. *Grundkurs theoretische Physik 5/2*. Springer Verlag, 2012.

[8] W. Nolting. *Grundkurs theoretische Physik 7*. Springer Verlag, 2005.

[9] U. Schollwöck, J. Richter, D. J. J. Farnelland, and R. F. Bishop. *Quantum Magnetism*. Springer Verlag, 2004.

[10] F. H. L. Essler, H. Frahm, F. Gohmann, A. Klumper, and V. E. Korepin Frontmatter. *The One-Dimensional Hubbard Model*. Cambridge University Press, 2005.

[11] W. Nolting. *Quantentheorie des Magnetismus - Teil 2: Modelle*. Teubner, 1986.

[12] J. Hubbard. Electron correlations in narrow energy bands. ii. the degenerate band case. *Proceedings of the Royal Society of London*, 1964.

[13] A. Altland and B. Simons. *Condensed Matter Field Theory*. Cambridge University Press, 2006.

[14] *Generalised Rayleigh Quotient Iteration for Lambda-Matrices, in: Archive for Rational Mechanics and Analysis, Vol. 8, Nr. 4.* Springer Verlag, 1961.

[15] W. Bunse. *Numerische Lineare Algebra*. Teubner, 1984.

[16] A. Meister. *Numerik linearer Gleichungssysteme*. Springer Verlag, 2011.

[17] W. Alt. *Nichtlineare Optimierung*. Springer Verlag, 2011.

[18] T. J. Rivlin. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*. Wiley, 1990.

[19] G. B. Arfken. *Mathematical Methods for Physicists*. Academic Press, 1985.

[20] G. Hämmerlin and K.-H. Hoffmann. *Numerische Mathematik*. Springer Verlag, 1994.

[21] L. N. Trefethen and D. Bau. *Numerical linear Algebra*. SIAM, 1997.

[22] Dirk Werner. *Funktionalanalysis*. Springer, 2006.

[23] Itai Arad, Alexei Kitaev, Zeph Landau, and Umesh Vazirani. An area law and sub-exponential algorithm for 1d systems. *arXiv: 1301.1162*, 2013.

[24] F. Verstraete and J. I. Cirac. Matrix product states representation ground states faithfully. *Physical Review B*, 2005.

[25] I. Peschel M. Kaulke and O. Legeza. Ann. physik (leipzig) 8. In *153 (1999)*.